2023

Online appendix for “Studying the interplay of party support and turnout; Chapter 5 in The Problem of Governing: essays for Richard Rose edited by Michael Keating, Ian McAllister, Edward Page and Guy Peters. London: Palgrave Macmillan

Mark N. Franklin

The published chapter for the Rose festschrift volume, **Studying the interplay of party support and turnout,** was simplified to save space but these appendices relate to the more comprehensive preprint, available in the same archive at https://digitalrepository.trincoll.edu/facpub/387/ They should be quite comprehensible when used with the published chapter.

CONTENTS of online appendices

ONLINE APPENDICES

A. Data

Respondent's left-right position, party's left-right position and which party respondents voted for were all coded at the individual respondent level in the Integrated Module Dataset (IMD) of the Comparative Study of Electoral Systems (Quinlen et al. 2018).

Questions giving rise to these measures were:

1. ["Which party did you vote for?"] (Question varies from country to country).

2. ["In politics people sometimes talk of left and right. Where would you place yourself on a scale from 0 to 10 where 0 means the left and 10 means the right?"] (Modules 1 and 2) "Where would you place yourself on this scale?" (Modules 3 and 4). Data recoded 0-1.

3. "Now, using the same scale, where would you place [Party A-F]?" (all modules).

To get a measure of left-right proximity I reshaped the individual-level data to the response level, where each response pertains to a separate party that respondents could vote for (or not) and for which they provide an estimate its left-right location. In political science we normally call this a stacked dataset, following Eijk et al. (2006). These stacked data were then collapsed (with values averaged across categories of variables defining the target level), either to the party level or to the party-birthyear level.[1]

Table A.1 displays univariate statistics, where they make sense, for variables employed in the main text at each level of aggregation used there. Note that generic party variables (variables having to do with parties in general ather than with specific named parties) and variables created from those

[1] Note that the party level of analysis subsumes the country level since each country has parties with unique codes – codes that are not repeated for parties of any other country. For respodents whose response was missing for any particular party, party left-right location at that level was plugged with the mean location across the non-missing responses of other respondents (chosen according to criteria set out in Appendix B3).

generic party variables, only exist in stacked (response-level) data and levels of analysis (party or birthyear cohort) derived from stacked data.

Table A.1 Univariate statistics for variables employed in the chapter and its appendices

	R's left-right location	Party 1 's left-right location	R voted for Party*	Generic party's left-right location	R's proximity to generic party	R's vote for generic party
Respondent level						
N of cases	120,015	116,409	114,207			
Minimum	0	0	360,001 *			
Maximum	10	10	8,400,004 *			
Mean	5.41	5.79				
Std deviation	2.44	2.99				
Response level (stacked)						
N of cases	779,790	667,791		667,791	725,229	926,895
Minimum	0	0		0	0	0
Maximum	10	10		10	10	1
Mean	5.40	5.03		5.03	7.13	0.12
Std deviation	2.46	2.95		2.95	2.48	0.32
Birthyear-cohort level						
N of cases	34,228	31,662 +		31,950	33,237	33,237
Minimum	0	0		0	0	0
Maximum	10	10		10	1	0.53
Mean	5.45	5.10		7.07	0.	0.06
Std deviation	1.17	2.06		1.18	0.08	0.08
Party level						
N of cases	526	++		493	493	503
Minimum	4.09			0.60	0	0
Maximum	6.98			9.02	1	0.53
Mean	5.44			5.02	0.65	0.11
Std deviation	0.58			1.81	0.16	0.12

Notes:* Party ID code. ** Measured at country-year level and duplicated onto birthyear/party levels.
+ N = 12.596 with appropriate lags. ++ N = 130 with appropriate lags.

Table A.2 lists all the elections conducted between 1996 and 2016 in each country that contributed at least 3 surveys to the CSES. Timepoints producing data for this paper are boldfaced. Australia, Israel and Japan each contributed four election studies to the IMD, but those were separated by additional elections rendering them non-contiguous, so no election studies from these countries are boldfaced. The final column counts the number of included studies.

Table A.2 Elections included in the span of time covered by the CSES IMD data, with boldfacing for adjacent elections yielding data included in analyses for the chapter and appendices

Sequence in analysis	1			2		3		4	5	Total included
Australia	1996	1998	2001	2004		2007	2010	2013		0
Canada	1997		2000	2004	2006	**2008**		**2011**	**2015**	3
Czech Republic	1996	1998		**2002**		**2006**		**2010**	**2013**	4
Germany	**1998**			**2002**		**2005**		**2009**	**2013**	5
Iceland	**1999**			**2003**		**2007**		**2009**	**2013**	5
Irelend	1997			**2002**		**2007**		**2011**		3
Israel	1996		1999	2003		2006	2009	2013		0
Japan	1996		2000	2004	2005	2007	2009	2013		0
Republic of Korea	**2000**			**2004**		**2008**		**2012**		4
Mexico	**1997**			**2000**		**2006**	2009	2012		3
New Zealand	1996			2002	2005	**2008**		**2011**	**2014**	3
Norway	**1997**			**2001**		**2005**		**2009**	**2013**	5
Peru	**2000**			**2001**		**2006**		**2011**	**2016**	5
Poland	**1997**			**2001**		**2005**		**2007**	**2011**	5
Romania	1996		2000	**2004**		**2009**		**2014**		3
Solvenia	1996		2000	**2004**		**2008**		**2011**		3
Sweden	**1998**			**2002**		**2006**	2010	2014		3
Switzerland	**1999**			**2003**		**2007**		**2011**		4
Taiwan	1996	1998	2001	**2004**		**2008**		**2012**		3
United States	1996		2000	**2004**		**2008**		**2012**		3

Bibliography

Quinlen, Steven, Christian Schimpf, Katharina Blinzler, Slaven Zivkovic, Bojan Todosijevic (2018) *The Comparative Study of Electoral Systems (CSES) Integrated Module Dataset (IMD)*. GESIS Leibniz Institute for the Social Sciences, Colone Germany. December 4, 2018 version. doi:10.7804/cses.imd.2018-12-04.

B. Robustness checks

B0 Number of lags

In the main text it was mentioned that error correction models can be employed with a variety of lag structures, by diagnosing the nature of any particular structure and then modeling the specific structure found. But diagnosing the structure is not always totally straightforward and, in footnote 8 in the main text I mention that the lag structure for Table 1's Model B is somewhat ambiguous, showing effects both from two elections prior and from three elections prior. The two versions of the model are shown in Table B0's Models B and B2, where the feedback variable is party support.

Table B0 Comparing (fixed) feedback effects on proximity and turnout for different lags (IMD data; Greek letter Δ labels each differenced variable $X_{t-1} – X_{t-3}$)

Level of analysis:	Birthyear cohort level		Birthyear cohort level	
	Model B	Model B2	Model C	Model C2
Outcome:	Δ.proximity	Δ.proximity	Δ.turnout	Δ.turnout
Feedback:	Party support	Party support	Proximity	Proximity
Inputs:	Coef. (s.e.)	Coef. (s.e.)	Coef. (s.e.)	Coef. (s.e.)
1) Lagged outcome (ECP)	-1.55 (0.14)	-1.38 (0.14)	-1.39 (0.02)	-1.42 (0.05)
2) Δ.Party support(log) †† $_t$			0.24 (0.02)	0.24 (0.01)
3) Party support(log) †† $_{t-1}$			0.43 (0.04)	0.37 (0.01)
4) Δ.Feedback variable $_{t-1}$		-0.29 (0.05)		0.22 (0.06)† ns
5) Feedback variable $_{t-2}$		-0.40 (0.10)		0.40 (0.11)† ns
6) Δ.Feedback variable $_{t-2}$	-0.30 (0.15) ns		-0.23 (0.14) ns	
7) Feedback variable $_{t-3}$	-0.64 (0.28)		-0.54 (0.21)	
Intercept	0.76 (0.10)	0.98 (0.02)	2.68 (0.21)	1.66 (0.09)
R-squared	0.80	0.67	0.91	0.83
Observations	73	918	920	920
Number of country-birthyrs	56	708	709	709

Note: All coefficients significant at $p<0.01$, one-tailed, unless marked ns (not significant).
† Model C2 feedback effects are not significant on one-tailed tests due to wrong signs.
†† The log transformations address artifacts due to very different distributions for the two inputs.

As can be seen, the coefficients for party support in Model B (from the main text) are considerably stronger that the coefficients shown in Model B2 that has the shorter lag (though the differences are not statistically significant at the 0.01 level) but based on many fewer observations than those in Model B2. The same is true for effects involving turnout, comparing Model C with Model C2, where the feedback variable is proximity to party. The theoretical expectations on which this research is based are better met with three-lag models but, with such models, the N is insufficient for appropriate tests to unambiguously rule out stationarity issues (see Appendix C).

B1 Level of analysis

A major concern for some scholars perusing this chapter might be the fact that analyses are conducted using data aggregated to a higher level of analysis (two different higher levels, in fact) than the level of aggregation at which the data were collected.[2] As explained in the main text, conceptually these are the levels of analysis relevant to the theorizing set out in the chapter; and few scholars will balk at seeing analyses relevant to understanding the behavior of political parties that are conducted at the party level of analysis, even if the data were originally collected at some other level of aggregation. Despite the fact that I can write the exact same phrase regarding birthyear cohorts, I know from bitter experience with journal reviewers that many of them do balk at passing on analyses conducted at the birthyear-cohort level. Generally they give no reason for doing so, treating the problem as self-evient; but some do mention the possibility that composition effects might threaten the findings. Indeed this is true, and also in regard to party-level analyses with aggregated data; but not because the aggregate-level findings are biased. Rather it would be because the individual-level data are adding non-random noise that would need to be totally controlled for if the aggregate-level findings are to be replicated with individual-level data. And totally controlling for individual-level effects is hard to do. It requires that EVERY individual-level variable correlated with the outcome of interest be known and included in the analysis, a virtually impossible task. The truth is that, for aggregate-level effects truly governed by aggregate-level processes, using data that has been aggregated to the appropriate level should be the safest approach, automatically removing whatever spurious effects would have threatened individual level findings. Put another way, analyses at the theoretically-defined level remove the need to control for spurious effects at the level of aggregation used for data collection.

[2] That level is the response level of aggregation, not the respondent level, since the questions at the center of my analysis were asked about each party separately. Answers originally occupied multiple variables for each respondent. When reshaped (stacked) each party-regarding variable became a separate case in the response-level dataset (Google search for “De Sio stackMe”).

However, differences between findings at the aggregate and individual levels are not sufficiently large for the superiority of cohort-level estimates to be demonstrated, as shown in Table B1a (for Models A and B of Table 1 in the main text).[3] What we see is rather that it does not matter to my findings whether I use party-level or birthyer-level data rather than response-level data (with or without controls for individual level turnout covariates). Effect coefficients are effectively the same with any of these estimation strategies.

Table B1a Comparing Table 1 Models A and B effects at party and birthyear level with response-level effects, where aggregate over time variables have been merged into the response-level data (Greek letter **Δ** prefixes differenced variables; for more details see footnote 3)

Origin for timevars:	Party level	Birthyear-level data		Birthyear-level data		
Level of analysis:	Model A Party level	Model A1 Response lvl w'out controls	Model A2 Response lvl with controls	Model B Birthyr lvl	Model B1 Responsel lvl w'out controls	Model B2 Response lvl with controls
Outcome:	**Δ.**Support	**Δ.**Support	**Δ.**Support	**Δ.**Proximity	**Δ.**Proximity	**Δ.**Proximity
Inputs:	Coef. (s.e.)	Coef. (s.e.)	Coef. (s.e.)	Coef. (s.e.)	Coef..(s.e.)	Coef. (s.e.)
1) Lagged outcome (ECP)	-1.07 (0.06)	-1.06 (0.00)	-1.24 (0.00)	-1.38 (0.14)	-1.29 (0.00)	-1.30 (0.00)
2) **Δ**.left-right proximity $_t$	0.23 (0.08)	0.27 (0.00)	0.28 (0.00)			
3) Left-right proximity $_{t-1}$	0.33 (0.12)	0.37 (0.00)	0.37 (0.00)			
6) **Δ**.support for party $_{t-2}$				-0.29 (0.05)	-0.22 (0.00)	-0.15 (0.00)
7) Support for party $_{t-3}$				-0.40 (0.10)	-0.30 (0.00)	-0.17 (0.01)
8) Individual level covariates †	NO	NO	YES	NO	NO	YES
Intercept	-0.07 (0.07)	0.11 (0.00)	0.06 (0.00)	0.98 (0.02)	1.07 (0.00)	0.81 (0.00)
R-squared	0.64	0.62	0.68	0.76	0.68	0.70
Observations	358	28,632	28,632	920	28,632	28,632
Number of cntry-birthyrs	167	709	709	709	709	709

Notes: All coefficients significant at p<0.001, one-tailed unless marked ns (not significant).

† Individual-level covariates are age, age^2, gender, married, religion, knowledge, efficacy, partisan, and satisfaction with democracy. Models A and B are taken from the main text.

[3] For both sets of analyses, missing responses at the individual level have been multiply-imputed, separately for each replication (context), using as predictors the other covariates employed in each model. Note that all variables in each analysis are individual-level versions except those for which past values play a part (this means that differenced variables other than the depvar and all lagged variables are averaged across parties or cohorts). The depvar is constructed by subtracting response-level turnout from turnout aggregated to the party-level (for Model A) or to the birthyear cohort-level (for other models) units at t-1.

Table B1b does the same thing for Model C of Table 1 in the main text, comparing findings using birthyear level with findings using response-level data.

Table B1b Comparing Table 1 Model C effects at birthyear level with response-level effects, where aggregate over time variables have been merged into the response-level data (Greek letter **Δ** prefixes differenced variables; for further details see footnote 3)

Origin for timevars:	Birthyear cohort data		
	Model C	Model C1	Model C2
Level of analysis:	Birthyr lvl	Response lvl without controls	Response lvl with controls
Outcome:	**Δ.**Turnout	**Δ.**Turnout	**Δ.**Turnout
Inputs:	Coef. (s.e.)	Coef. (s.e.)	Coef. (s.e.)
1) Lagged outcome (ECP)	-1.39 (0.02)	-1.68 (0.01)	-1.63 (0.00)
2) **Δ**Support for party(log)$_t$	0.24 (0.02)	0.23 (0.00)	0.32 (0.00)
3) Support for party(log) $_{t-1}$	0.43 (0.07)	0.41 (0.00)	0.52 (0.00)
4) **Δ**.left-right proximity $_{t-2}$	-0.23 (0.14)ns	-0.26 (0.01)	-0.00 (0.01)ns
5) Left-right proximity $_{t-3}$	-0.54 (0.21)	-0.69 (0.01)	-0.35 (0.02)
6) Individual level covariates†	NO	NO	YES
Intercept	2.68 (0.21)	2.72 (0.02)	1.95 (0.02)
R-squared	0.91	0.92	0.93
Observations	920	28,632	28,632
Number of cntry-birthyrs	709	709	709

Notes: All coefficients significant at $p<0.01$, one-tailed, unless marked ns (not significant).
† Individual-level covariates are age, age^2, gender, education, married, religion, income, urban, union, knowledge, efficacy, partisanship, and satisfaction with democracy. Model C s taken from the main text.

The contrasts that we see in these two tables regarding effects measured at different levels of analysis are pretty inconclusive as concerns the superiority or otherwise of estimations made at the level at which a variable was theorized to have its effects. Individual-level data into which time-serial indicators at a higher level of analysis have been merged produce effects that are largely indistinguishable from the effects at the level at which the time serial measures were obtained. Note that the variance explained in all these models is individual-level variance (actually, response-level variance since the individual-level data were reshaped to the level of parties within respondents).

B2 Do turnout findings at birthyear-cohort level reflect actual turnout evolution?

Meanwhile we need to address a question glossed over in the main text for lack of space: whether birthyear cohort findings regarding the knock-on effects of feedback for party support actually correspond to meaningful effects on turnout when all birthyear cohorts present for a particular election are taken together as a single case. This question is easily addressed by using model C from Table 1 in the main text to predict differenced turnout at the birthyear cohort level. I then aggregate the data (including the newly predicted values) to the country-election level and estimate differenced turnout at that level from the two sets of predicted values. Results are in Table B2.

As can be seen, Model B2a has only one coefficient that is statistically significant. The lack of significance for the coefficients that would make this an error correction model suggests that it should not be seen as such but rather as a straightforward regression model. Such a model is presented in Model B2b. The single significant coefficient in this analysis,[4] explaining 73 percent of variance, strongly supports the supposition made in the main text that synchronization across parties would ensure effects on election-level turnout. Evidently this topic needs further attention.

Table B2 Country-election level turnout explained by predicted turnout from birthyear turnout, estimated by Model C of Table 1 in the main text

Outcome:	Model B2a Differenced turnout	Model B2b Differenced turnout
Inputs:	Coef. (s.e.)	Coef. (s.e.)
1) Lagged turnout	-0.85 (1.27)ns	
2) Δ.turnout predicted by Model C2 $_{(t-1)}$	1.92 (0.86)	1.54(0.15)
3) Δ.turnout predicted by Model D2 $_{(t-2)}$	1.46 (0.35)ns	
Intercept	0.62 (0.97)ns	-0.01(0.01)ns
R-squared	0.78	0.73
Observations (separate elections)	42	70
Number of countries	26	28

Note: Coefficients significant at $p<0.05$, one-tailed, in Model B2a; 0.001, one-tailed, in Model B2b, unless marked ns (not significant).

[4] The limited N results from degrees of freedom used up by lagged terms. The coefficient suggests that actual turnout increases by about 1.5 percent for every 1 percent predicted at the birthyear cohort level. In turn, at that level, Table 1 findings (repeated in Table B1b above) suggest about half a point change in turnout for each one-point change in party support, where units are relative to maximum change in support. So our models predict about 75 percent of the empirical range of turnout change, leaving 25 percent for other forces.

B3 Endogeneity problems when estimating persuasion and learning effects

As pointed out in the main text, attempting to estimate the relative contribution of each component of a proximity measure, as I do in Table 2 in the main text, evidently yields grave risks of findings contaminated by endogeneity. And, as mentioned there, the method chosen to create the proximity measures used in that table is virtually the only one available that does not show endogeneity artifacts. Here we elaborate on that assertion.

Left-right proximity measures are constructed by taking the absolute value of the difference between measures of party location and of respondent location. Measurement of self-assessed respondent locations are disussed in the main text, footnote 11. Here we address problems found in measures of party location. The measure originally employed by Franklin and Lutz (2020), the protopype for the research reported here, is straightforward. Principle Investigators for each survey were asked to code their country's parties appropriately and, from this measure along with respondent self-assessed location, measures of proximity were constructed. However, while widely used in cognate research, such a measure is not the one we really want when we are studying changes in party location. This is because the experts doing the judging are not necessarily very quick to pick up on changes in party location, as we shall see.

The most common alternative to expert-judged party locations is to employ respondent-judged party locations. But when respondent judgements are used in place of expert judgemens, the question arises what to do about missing judgements? The most widely employed solution is to "plug" the missing party locations with the average value assigned by respondents who answered the party location question. But if some respondents are positioning a party on the basis of projection then the plugging value will reflect the most widespread projection effect – probably a bias towards the largest party (the same reasoning applies to assimilation effects). So an alternative strategy when finding an average plugging value is to ignore judgements that are the same as the respondent's own self-evaluated left-right position. These will be referred to in what follows as "difference-plugged" party locations, distinguishing them from "all-plugged" locations where means are based on judgements derived from all non-missing responses. Diff-plugging reduces the number of respondents responsible for the party placements but increases the validity of the responses obtained.

Table B3a compares results of analyses such as those presented in Table 2 of the main text when proximities based on expert-assessed party locations are compared with proximities based on all-plugged and diff-plugged respondent-assessed locations. Model A shows that, with findings based on expert assessments (rows 2 and 3), parties play no role in maintaining left-right congruence, presumably because principal investigators are largely the same individuals from election to election and may not very quickly revise their opinions about where parties stand. So this measure would not serve us well when investigating party responsiveness to respondent issue concerns.

Table B3a Change in left-right proximity due to change in respondent vs party left-right location (birthyear cohort analyses with expert vs respondent party placements)

	Model A	Model B	Model C
		All-plugged	Diff-plugged
Outcome:	Expert-assessed	Resp-assessed	Resp-assessed
	Δ.Proximity	**Δ.**Proximity	**Δ.**Proximity
Inputs:	Coef. (s.e.)	Coef. (s.e.)	Coef. (s.e.)
1) Lagged outcome (ECP)	-1.27 (0.02)	-1.58 (0.02)	-1.32 (0.02)
2) **Δ.**Party l-r location $_{t}$	-0.01 (0.02)ns	-0.26 (0.04)	0.23 (0.03)
3) Party left-right location $_{t-1}$	0.01 (0.02)ns	-0.25 (0.06)	0.33 (0.04)
4) **Δ.**Supporter l-r location $_{t}$	0.06 (0.02)	0.14 (0.02)	-0.07 (0.01)
5) Supporter l-r location $_{t-1}$	0.13 (0.03)	0.33 (0.03)	-0.05 (0.02)*
6) Constant	0.68 (0.02)	0.30 (0.04)	0.77 (0.03)
R-squared	0.66	0.72	0.65
Observations	4,357	4,469	4,544
Country-birthyears	1,961	1,962	1,971

Notes: All coefficients significant at the p<0.01 level, one-tailed, except as marked, with "*" for significance at the 0.05 level. Note that expert assessments pertain only to parties.

Endogeneity is also a concern. Model B shows such effects most obviously. It explains most variance of the three models and shows a larger long-term contribution from respondent shifts in left-right location than from party shifts, contrary to expectations of party responsiveness. That coefficient (of 0.33, row 5) includes endogeneity by construction, since effects take account of party positions assigned by respondents who place the parties where they place themselves.

We next come to the measurement strategy actually employed in the main text. What roles do we find for party and respondent contributions to the maintenance of proximity when endogeneity is removed? Model C replicates the analysis presented in Model B of Table 2 in the main text, which rules out projection by ignoring positions assigned by those respondents who place the parties where they place themselves.[5] This "lobotomization" of voter assessments certainly removes any possible projection or assimilation effects, but it will also have eliminated theorized learning effects for voters who placed themselves where they observed their favored party to be positioned. So it conducts a very stringent test for supporter influence, which is still passed – even if only at the p=0.05 level of statistical significance. True effects of party supporters on measured proximity in voting models must lie somewhere between the absolute values of effects shown in Models A and C (quite small in any case). For turnout models we can be a bit more specific.

Turning now to the persuasion/learning effects that are the primary source we expect for feedback in turnout models, we use as our laboratory one of the analyses reported in Table B0 of this appendix. Table B3b starts by repeating (in Model D) Model C2 of Table B0 (the two-lag version of Model C from the main text's Table 1). We use this model so as to have sufficient N for the experiment contained in Models G and H below. In Table B3b we adapt Table B0's Model C2 by replacing measures of proximity with measures of left-right location underlying those proximities. Successive models progressively adapt Model D, first by removing the measure of supporter left-right location to demonstrate that it adds nothing to R-squared (in Model E) and then by changing the difference-plugged respondent-assessed measure of left-right party location to an all-plugged measure in Model G. We see that the two measures produce identical coefficients when statistically significant.[6] In

[5] Model C also switches the signs of effects to accord with the dominant influence found (see footnote 13 in the main text). Note that none of these models were evaluated by Franklin and Lutz (2020) who focused uniquely on party-level analyses where the deficiencies of expert-ratings were not apparent.

[6] This might seem rather strange, given the marked differences between findings for the same comparison in

Models G and H the measure employed is the same (expert-assessed) measure that was used in Model A. Model G makes it clear why expert assessed measures were not used in the main text: its robust effects have the wrong signs. Clearly the findings in earlier models are due in part to effects not present in expert assessments.

Table B3b Replications of an Appendix B0 version of Table 1, Model C, effects on turnout using various measures of respondent-party left-right proximity

Outcome: **Δ.**Turnout Inputs:	Model D Diff-plugged resp-assessed Coef. (s.e.)	Model E Diff-plugged resp-assessed Coef. (s.e.)	Model F All-plugged resp-assessed Coef. (s.e.)	Model G Exprt assessd second lag Coef. (s.e.)	Model H Exprt assessd third lag Coef. (s.e.)
1) Lagged outcome (ECP)	-1.47 (0.03)	-1.45 (0.03)	-1.45 (0.03)	-1.48 (0.03)	-1.64 (0.08)
2) **Δ.**Support for party (log) $_t$	0.25 (0.01)	0.25 (0.01)	0.25 (0.01)	0.25 (0.01)	0.24 (0.02)
3) Support for party (log) $_{t-1}$	0.39 (0.02)	0.38 (0.02)	0.38 (0.01)	0.40 (0.02)	0.37 (0.04)
4) **Δ.**Party l-r location $_{t-1}$	0.11 (0.08)ns	0.19 (0.07)ns	0.16 (0.07)ns	0.19 (0.03)ns	-0.63 (0.48)ns
5) Party left-right location $_{t-2}$	-0.38 (0.12)	-0.24 (0.11)	-0.24 (0.11)	0.21 (0.04)ns	-1.04 (0.94)ns
6) **Δ.**Supporter l-r location $_{t-1}$	-0.11 (0.04)				
7) Supporter l-r location $_{t-2}$	-0.16 (0.06)				
6) Constant	2.27 (0.08)	2.09 (0.06)	2.07 (0.06)	1.95 (0.05)	2.62 (0.43)
R-squared	0.85	0.85	0.85	0.85	0.92
Observations	2,612	2,629	2,629	2,453	872
Country-birthyears	1,700	1,711	1,711	1,582	708

Notes: All coefficients significant at the $p<0.01$ level, one-tailed, except as marked "ns". Effects of party left-right location in Model G are not significant, one tailed, because they have the wrong signs.

If we were to make a list of reasons why expert-assessed party positions would fail to track (or to be tracked by) respondent positions, such a list would clearly not include projection, assimilation or learning, all of which would strengthen observed effects due to endogeneity with respondent locations. Instead, these effects (if we accept the evidence of Models D and E that the effects are real)[7] are likely being suppressed in Model G by experts whose judgements regarding party locations

---

Table B3a. But in Table B3a the comparison is between models using different dependent variables whereas here the comparison is between use of different independent variables; and the indep that receives the quasi-experimental treatnent is precisely the one that has a different coefficient in Model F than in Model E.

[7] Any projection/assimilation effects would cause coefficients in Model F to differ from those in Model E, since expected bias from respondents whose left-right locations echoed the locations of parties they placed

do not change over time even though the parties' locations do in fact change. Evidently this could not happen with repeated samples of survey respondents but is quite possible if the same PIs are engaged in successive election studies, as already mentioned. Model H supports this supposition by showing correctly-signed effects (even if not statistically significant) when an additional lag provides more time for PI's judgements to evolve. So party positioning has real and quite marked effects on the decision to vote, approaching 0.4 (the strongest substantive effect in the table). This would be quite remarkable, given that parties do not vote, were it not for my supposition that what we are seeing are effects of changing party positions transmitted to newly adult citizens during their young adult socialization process and helping to determine their decisions to vote or not.

Considerations developed in this section of Appendix B also support my choice of proximity measure to employ in the main text.[8] However, these considerations might still be considered suspect in the absence of confirmatory findings from survey experiments that we turn to next.

B4 Validating voter awareness of policy shifts

Adams et al. (2018) stress the need to validate evidence for voter awareness of policy shifts by verifying that the evidence remains compelling when cognate variables are employed. Those authors might have added that it would also be important to confirm the findings using alternative data sources. In this part of Appendix B I do both, validating my evidence of voter awareness (of shifts in perceived left-right proximity to the parties they support) by employing an alternative measure of awareness – the reports those voters make of liking a party – and, additionally, referencing findings from an alternative data source.

---

should have been eliminated in diff-plugged values. Yet, in practice, differences between these two models are virtually absent, validating their measured effects as real.

[8] They also provide an informal test for an implication implicit in my theorizing that short-term ECM effects will not be widespread because they do not invariably result from the sort of exogenous shocks that, in many applications, lead to disequilibria in need of correction. In the tables of this section long-term (differenced) effects are generally smaller than short-term (lagged) effects and often not statistically significant.

Table B4 focuses on responsiveness, with either proximity and liking as alternative outcomes. The first model replicates the 3 lag model shown in Model B of Table 1 in the main text, already replicated as Model B of Table B0 in this appendix; the second model replicates the 2 lag model shown as Model B2 of Table B0. The other two models repeat the first two while using a ten-point likes/dislikes measure in place of the ten-point proximity measure.

As can be seen, effects of interest for the three-lag version of the likes/dislikes replication (the version with lags that match those employed in the main text) is not statistically significant, but all effects in the two-lag model (Model B2a) show a high level of significance. The question asked in Section B0 regarding number of lags would have to be re-assessed if we were to use the party likes measure to replace party proximity; but doing so is not at issue. The purpose of these madels is simply to show that respondents react to a similar measure in much the same way as they react to the proximity measure, suggesting voter awareness of changes in party stances.

Table B4 Robustness of representation across measures of birthyear level voter awareness

	Model B	Model B2	Model Ba	Model B2a
Outcome:	Party proximity	Party proximity	Party liking	Party liking
Inputs:	Coef. (s.e.)	Coef. (s.e.)	Coef. (s.e.)	Coef. (s.e.)
Outcome $_{t-1}$	-1.55 (0.14)	-1.38 (0.03)	-1.59 (0.06)	-1.20 (0.03)
Differenced support $_{t-1}$		-0.29 (0.05)		-0.14 (0.05)
Support $_{t-2}$		-0.40 (0.10)		-0.50 (0.02)
Differenced support $_{t-2}$	-0.30 (0.15)ns		-0.09 (0.16)ns	
Support $_{t-3}$	-0.64 (0.28)		-0.20 (0.27)ns	
Constant	0.76 (0.10)	0.98 (0.02)	0.72 (0.04)	0.57 (0.02)
R-squared	0.80	0.67	0.81	0.64
Observations	73	2 ,577	935	2,670
Number of groups	56	1,660	724	1,738

Note: All coefficients significant at p<0.001 unless marked “ns”.

In cognate research, the same strategy yielded substantively identical findings regarding long-term equilibria for party support and left-right proximity based on a completely different data source (European Parliament election studies) and a time-period (1989-2014) that was somewhat longer (Franklin 2015).

C. Error correction models, stationarity and co-integration

The findings of this paper are produced by error correction models (ECMs). For such models to yield valid findings a number of requirements must be met, of which the primary one is also the most difficult to verify: the process under investigation must be in long-term equilibrium (so any short-run disequilibrium will be corrected in due course).[9] This means that, over the long term, all dynamic elements in the model must either be stationary or else co-integrated with the relevant element(s) on the other side of the equals sign (De Boef and Keele 2008). Stationarity means simply a long-run stable mean to which a series reverts after any deviation while co-integration means two series moving together (thus both non-stationary) in a long-run relationship such that the linear combination of the two series is stationary.

Confirming stationarity or co-integration with regard to my data is not straightforward, however. My time-series for specific country panels are very short: no more than five time-points with an average of 3.5; and I have only 57 parties with data for 4 or more time-points. Given random perturbations in level of support and left-right position of individual parties, a sample of just three or four cases can be expected to show trends that are upward, downward or both, pretty much at random. Examination of my data confirms this expectation at levels of statistical significance appropriate to the small Ns involved. Stationarity is found by an Augmented Dickey-Fuller unit-root test at the 0.05 level for 7 (or 8) out of 69 (or 72) parties (thus random at virtually the 0.1 level of statisticl significance) while a Westerlund test finds co-integration at the 0.05 level in a further 13 out of 57 (thus random at close to the 0.2 level).

I take a two-pronged approach to arguing that my findings are not vitiated by what appear to be short-term anomalies. The first prong is to assert that, from a theoretical standpoint, my data should be in equilibrium and any indication to the contrary is thus spurious – simply capitalizing on

9 A standard ECM, by construction, meets requirements for balance, a major concern for contributors to a symposium on the topic (Keele, Linn and Webb 2016). There are differences of opinion regarding other requirements but in this appendix I apply the most stringent of the various possibilities.

chance variations evident in the short-term. This expectation receives face validity for my party-level analyses because the disequilibriating short-term effects in those analyses are not remotely statistically significant. For the birthyear cohort analyses, however, the disequilibriating long-term effects, though small (as expected), are nevertheless statistially significant. The second, more demanding, approach is to take indications of non-cointigrated non-stationarity (in the case of any given panel) at face value and demonstrate that when I exclude such panels my findings are substantively unchanged.

The starting point for any assessment of either co-integration or stationarity must be theoretical. I have no basis for supposing that the variables in my models would be co-integrated (which implies non-stationarity) because I do have every reason to suppose that they should be stationary. Political parties come and go, but those that came or went are not part of my sample, which consists only of parties present in the data for at least three time-points. Recent research has established that such parties tend to receive a level of support that is in a long-run stationary equilibrium, (Weber and Franklin 2018). Much the same applies to the left-right locations of such parties (Dalton and McAllister 2015). So I expect stationarity for my primary variables.

Of course such long-term equilibria are quite consistent with the appearance of short-term dis-equilibria in specific panels, simply on the basis of random perturbations (as already mentioned); and my data do fail a test of the "joint requirement" that both of my primary variables (left-right proximity and party support) are co-integrated (using a Westerlund test) in any panel that lacks stationarity for one or both of these variables (according to an Augmented Dickey-Fuller unit-root test). However, as already mentioned, the relatively small number of failures to meet the joint requirement is quite consistent with the notion that these failures are just random purturbations, only to be expected when taking short-term "snap shots" of data that, over a longer term, would have proved stationary.

Still, if I proceed with the second prong of my approach by taking at face value the individual failures to meet the joint requirement mentioned above, I can select for analysis just the panels that are "clean" in the sense that my primary variables are either stationary or cointegrated. If my results are substantively unchanged when using clean data this will lend support to the idea that, even though the joint requirement cannot be shown to have been met, still my findings are not artifacts of any departure from the clean data requirement.

Table C shows the results.[10] Although significance levels are sometimes low because of

Table C Feedback and representation in data selected for demonstrable stationarity or co-integration at p<0.05 (same models as Table 1 in main text, but smaller N)

Outcome variable:	Model A Differenced party support (party level)	Model B Differencd proximity (party level)	Model C Turnout (birthyr lvl)
Inputs:	Coeff (s.e.)	Coeff (s.e.)	Coeff (s.e.)
1) Lagged depvar	-1.21 (0.10)	-0.91 (0.27)	-1.61 (0.13)
2) Differenced party support	0.34 (0.16)		0.23 (0.04)
3) Party support $_{t-1}$	0.37 (0.29)*		0.40 (0.09)
5) Differenced party support $_{t-2}$		-0.12 (0.07)*	-0.91 (0.31)
6) Party support $_{t-3}$		-0.20 (0.12)*	-1.02 (0.44)
7) Constant	0.72 (0.07)	0.13 (0.24)	3.50 (0.47)
R-squared	0.76	0.79	0.98
Observations	83	26	203
Number of parties/birthyear chrts	31	20	181

Notes: Fixed effects regression analysis with standard errors in parentheses.
All coefficients significant at 0.05, one-tailed, unless marked "ns" (not significant) or * (significant at the 0.1 level, one-tailed).

the small N selected for analysis, and the outrageously high R2 for Model C suggests the possibility of over-fitting (Keele et al. 2016), the table shows coefficients for the various lagged terms that do not contradict those in Table 1 in the main text, though the feedback coefficients for Model C are implausibly large (but not large enough to call for the log transformation used in the main text).

[10] At both the party level and birthyear-cohort level, the resulting selection of panels is strongly enough balanced to be tested for stationarity of all remaining panels, and stationarity cannot be rejecteded at the 0.1 level either for proximity or for party support. Although my earlier tests for stationarity used the 0.05 level of significance, as already mentioned, I cannot apply that level to this prong of my investigation because too few cases in the party-level dataset survive selection by that criterion for the models I want to test.

Models with only two lags for feedback (not shown here but see Appendix B) are more convincing, raising again the question of whether I should be using two lags or three. But that question is a tricky one. The removal of countries that contributed non-stationary components to the pooled time-series may have become more telling with the reduction in available N due to the extra lag. So the question cannot be answered in the absence of more lengthy time series and so must wait on future research.[11]

Bibliography

Dalton, R., D. Farrell and I. McAllister (2011) *Political Parties and Democratic Linkage: How Parties Organize Democracy*. Oxford: Oxford University Press.

Franklin, M. (2015) "Dealignment and Accountability." Paper presented at the ECPR Joint Sessions Workshop on "Accountability without parties" in Warsaw, Poland, April. (Available on researchgate.net by means of a Google Scholar search).

Grant, T and M. Lebo (2016) "Error Correction Methods with Political Time Series" *Political Analysis* 24(1), 3-30.

Keele, L., Linn, S., & Webb, C. (2016). "Treating Time with All Due Seriousness." *Political Analysis, 24*(1), 31-41. doi:10.1093/pan/mpv031

Weber, T. and M. Franklin (2018) "A behavioral theory of electoral structure." *Political Behavior* (40): 831–856.

[11] If I compare the 2-lag results from clean data with two-lag results from the full dataset, that includes questionable panels, coefficients from the full dataset are generally more highly significant, statistically (coming from a larger dataset) but those coefficients are not statistically distinguishable from coefficients deriving from the partial set of demonstrably clean data. With three lags the disparity is greater but the coefficients from the partial dataset, though not statistically significant themselves, are still not distinguishable, statistically, from the significant coefficients in the full dataset. More importantly, the findings with two lags suggest that all panels in the full dataset are stationary or (when not stationary) co-integrated, a finding that should apply as much to effects from the third lag as to other effects.

D. What is to be done?

This is an unusual appendix; perhaps unique.

The substantive content of the chapter to which it belongs is the result of an accident. While testing for negative feedback in party support I realized that my measure of support at the party level (votes cast for each party as a proportion of electorate size) was indistinguishable, at the party level of analysis, from a measure of turnout. So it occurred to me to wonder what would happen if I substituted an actual measure of turnout (which would take on the same value for all of the parties competing in each specific election) for my measure of party support in the party-level dataset. The rest, as they say, is history and gave rise to the Rose Festschrift chapter and its online appendices A to C.

The analyses presented in the chapter and its first three appendices are first cuts at an account of why we would expect voter turnout to respond thermostatically to party support and policy congruence. Although the chapter bravely presents the findings as meeting conventional criteria for statistically significant findings, many alternative model choices would have been possible. And, although the robustness checks in Appendices B and C are supportive, many additional robustness checks are surely called for.

More importantly, the paper is unclear as to where it places itself in academic perspective. What we have here might be seen as a proposal for a new subfield in electoral research – a subfield that builds on the original promise of *The American Voter* (1960) to address both voting choice and turnout in concert, using linked theoretical foundations and analytic tools. However, the chapter might equally be seen as belonging with work on thermostatic governance. Knowing where to place the chapter in scholarly terms is critical to deciding how to frame it for maximum impact.

Given this uncertainty, what we have with this chapter should be seen as an academic doodle rather than a serious contribution. It has not been subjected to more than rudimentary peer review and I am fairly sure that submitting the piece, as written, for review by a major journal would result in reviews that would be critical (perhaps outraged) by the extent of the leap that the paper takes, beyond what are the contemporary frontiers on research into negative feedback in policy-making and on research into equilibriating processes in turnout and party support. The fact that the paper might be seen as addressing at least three separate subfields would probably guarantee outrage rather than just criticism.

Ignoring a quite large literature assessing the effects of turnout on the electoral fortunes of specific parties or party types, I seem to be the only person to have addressed this research question as such in over 60 years; and it is clear that I am in way beyond my depth. I have bitten off far more than I can chew. I need help.

I need help just to decide what should be the first step in any serious attempt to bring the ideas presented here into the mainstream of electoral research; and I am very open to the possibility that I have completely missed a major snag hidden somewhere in my own apparent findings. This is why, when asked to organize a Round Table on any topic of my choice, I proposed to organize one on this topic.[12] I hope that, if a sufficient number of scholars who are cleverer than me put their heads together, they may either bring this proto-project to a definitive close or else come up with a viable plan for moving it forward.

[12] See the European Academy of Sciences and Arts Call for Papers: *It's About People* (https://conference.almamater.si/the-2024-call/).