Web Appendix for "Levels of Linkage: Across-Agreement v. Within-Agreement Explanations of Consensus Formation Among States"

Abstract

In this web appendix, we expand on the nature of the substantive interview data on which we drew in the paper, as well as to show the robustness of the statistical results presented. The first part of the appendix describes, in detail, the nature of the interview evidence that was gathered for this project. The second part presents multiple alternative statistical models, and the effects associated with across-agreement and within-agreement linkage in these different models.

Elaboration of Interview Evidence

In this web appendix, we expand on the nature of the substantive interview data on which we drew in the paper, as well as to show the robustness of the statistical results presented. The first part of the appendix describes, in detail, the nature of the interview evidence that was gathered for this project. The second part presents multiple alternative statistical models, and the effects associated with across-agreement and within-agreement linkage in these different models.

Specifically, 146 interviews were conducted with state representatives involved in the EU decision-making process. These interviews were distributed across the EU member states as follows: Austria (4), Belgium (5), Bulgaria (1), Cyprus (3), Czech Republic (3), Denmark (8), Estonia (5), Finland (4), France (4), Germany (6), Greece (4), Hungary (3), Ireland (4), Italy (8), Latvia (3), Lithuania (5), Luxembourg (6), Malta (0), Netherlands (9), Poland (3), Portugal (5), Romania (4), Slovakia (1), Slovenia (3), Spain (0), Sweden (8), and United Kingdom (5). The remaining interviews were conducted with members of the European Commission and Council Secretariat involved with specific case negotiations that were analyzed. These individuals were present in the member state negotiations of those cases, but were not representing member state interests. No major player in the EU is therefore left out, and no single state dominates the interview process.

The interviews were also coded, regarding their analysis of across-agreement and within-agreement literature. The main patterns are highlighted here. Two types of questions were asked in the interviews. The first set of questions focused on the negotiation of a variety of specific dossiers in which the interviewee was a participant, and the specific strategies he/she and other member state representatives used in those negotiations as they pushed forward their states' interests, while at the same time working to reach a cooperative agreement. A total of 25 dossiers were covered, in detail, with perspectives on the negotiation of each from a wide variety of member states. A second set of questions focused on the more general level of analysis, asking questions about patterns, and the general types of strategies negotiators use when trying to forge an agreement in the institutions of

the Council of Ministers (from the Working Group and COREPER I and II, to the Council).

In the discussions of specific dossier negotiations, 93% of interviewees described characteristics of issue linkage as one of the strategies used to help bring about an agreement. These characteristics could include a specific tit-for-tat trade among two or more states/coalitions across specific issues/Articles of the dossier as well as a general discussion of reciprocity across issues with some states pushing on particular issues and "being flexible" or "giving" on others, and vice versa. Only 16% of interviewees highlighted any type of linkage to an outside dossier (or any other outside issue or political discussion) as being a key part of the process of forging an agreement in a particular dossier they were discussing. Moreover, 44% of these logrolling codings related simply to the linkage between the negotiation of the 2007-2013 Financial Perspective and the negotiation of the 7th Framework Research Programme and Competitiveness and Innovation Programme, as the Financial Perspective would lay out the budget for the next 5 years, and thus which directly affected the EU's research budget, and thus its ability to fund these programmes.

In the general questions, representatives were asked what types of strategies, in general, they observed being used to help forge agreements. Some representatives did mention logrolling and issue linkage as two types of such potential strategies. If they did not specifically mention these strategies, and time remained, follow-up questions were posed about these two specific strategies. No questions were posed about frequency – all claims about frequency were made, without prompt, by the interviewee.

Out of all those who were posed questions about logrolling (or who raised it as a strategy themselves), 77% added (on their own) that the use of such a strategy was "not the norm", "relatively rare", "rarely occurred", etc. Of the representatives interviewed, only 38% actually brought up logrolling as a strategy on their own. Moreover, of those representatives, in particular, a significant percentage (80%) still added that while logrolling was a strategy that was sometimes used, it was a rare occurrence.

No patterns existed in the types of states from which representatives came and their discussion of logrolling strategies. Some representatives from both small and large states, both new (i.e., joined in the 2004 or 2007 expansions of the EU) and old, mentioned logrolling while some from all four types of states did not. Those who explicitly brought up logrolling as being rarely used also spanned all four categories, as did the states that did not. The only pattern that arose regarding the type of state, was that representatives who discussed logrolling, when they were explicit, most often described logrolling as occurring (when it occurred) mostly between/among the "big" states. However, representatives from small, large, new and old states made this observation.

Statistical Robustness Checks

Results Robust to Imputation of Missing Data

This section of the robustness checks is to demonstrate that the results reported in the paper are robust to alternative imputation specifications for missing data. In particular, when coding the within-agreement linkage measure, missing data on position and salience for the EU states were imputed using Amelia (King, et al. 2001). Overall the imputed data are a small component of our main independent variable. We are only missing data for 8% of the data on states' positions on the various within-agreement issues, and only 7% of the data on the salience of those issues. Moreover, the missing data tends to occur in the same observation. Indeed, in 6.6% of cases we are missing both position and salience. The missing data for the *Within-agreement linkage* measure, overall, is therefore not a significant proportion of the data needed to create our Within-agreement linkage measure.

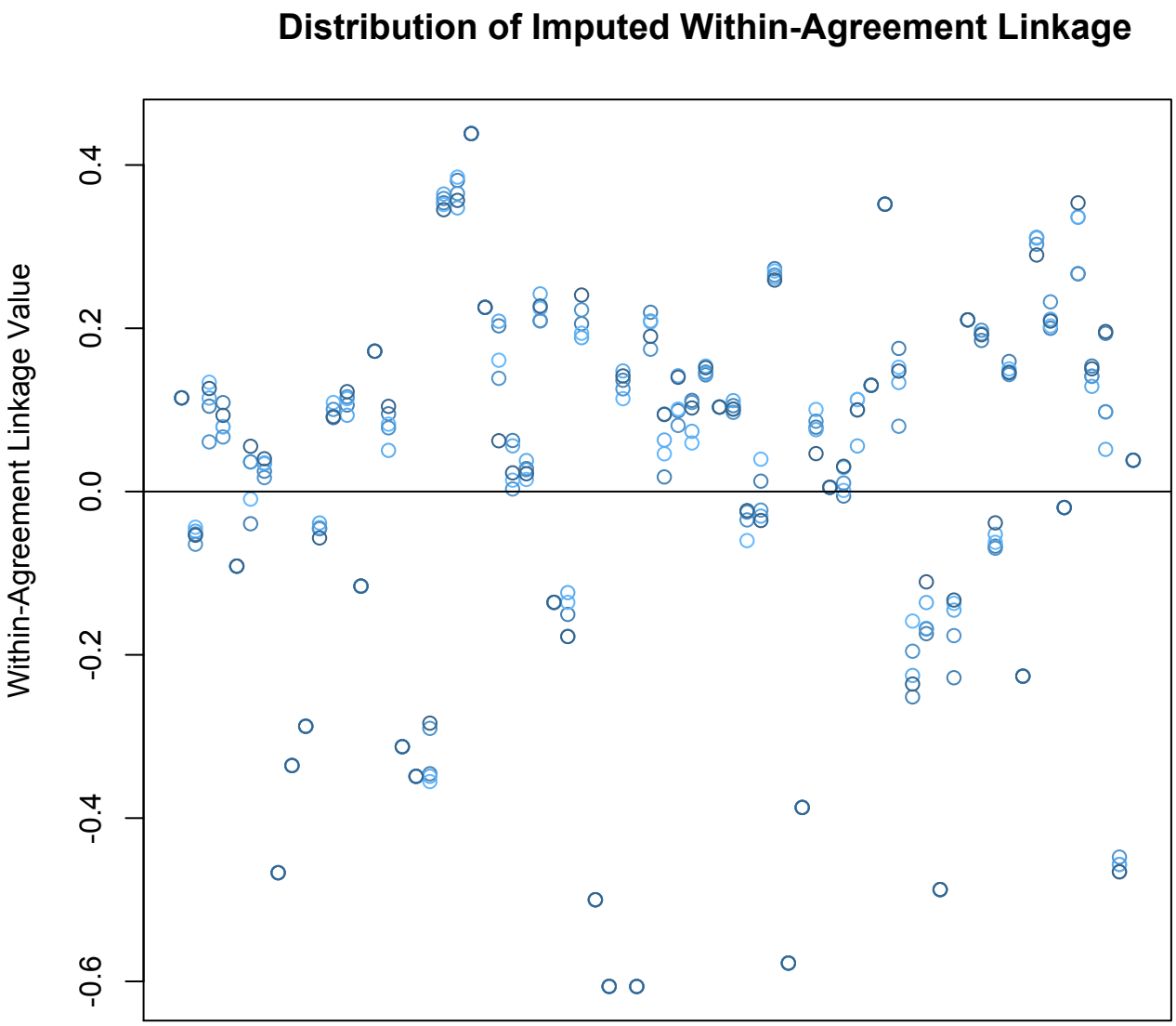
However, we still wanted to ensure that the results of the analysis were robust to to the imputation process. This was done in several ways. First, we take into account the fact that there might be different *reasons* for the data to be missing. Data might be missing, first, because an issue may not have been salient enough for a country to have taken a position, and so the data *appear* to be missing, but really just do not exist. Alternatively, a state's position and/or salience on a particular issue may simply not have been observed – i.e., the data really are missing. When imputing, we take into account whether the data are simply missing or actually do not exist.

We therefore code three alternative measures to account for these different reasons for missingness. The **Main** measure assumed that if both the position and salience for a particular state are missing, it is likely that the issue was not salient enough for a country to have taken a position, and therefore the data do not exist. If only one or the other is missing, that data is imputed. **Alternative Specification 1** treats all missing data as missing, and thus imputes values for all missing data on position and salience. Finally, *alternative specification 2* sets the salience of an issue to 0 and then imputes a state's position when both position and salience are missing. When only one or the other was missing, that value is imputed, as normal.

For each of these three measures, five alternative values of the missing data are calculated for each negotiation, resulting in five alternative values of the *within-agreement linkage* measure. Figure 1 plots the values of each of these five values for each negotiation, showing how these values tend to cluster together, in spite of the imputation process. For each negotiation, the values all tend to cluster together above zero (indicating a set of issues that are differently valued) or below zero (indicating a set of issues that are not differently valued). For every few negotiations do the five values ever cross the zero-threshold. The imputation process itself, which fills in only the missing observations from the data, therefore does not result in significant changes in the value of this key variable.

We then run the Bayesian analysis using these three alternative measures. **Model A** in **Table 1**presents the results using the main imputation specification – and the results thus parallel those presented in Model 1 in the paper. **Model**

Figure 1: Distribution of Five Imputed Within-Agreement Values for Each Proposed Agreement



B in **Table 2** reports the results derived from re-running Model A with **alternative specification 1** for conceptualizing and imputing the missing data. The results show that over 95% of the time the draws from the posterior from the within-agreement linkage measurement are greater than 0, indicating a positive

relationship between that variable and consensus. This leads us to conclude that under alternative specification 1, the same positive relationship exists as the one provided in our original measurement. We do not find such a strong relationship between consensus and the across-agreement measurement; the distribution is too wide to make any definitive claims. Finally, **Model C** in **Table 3** reports the results derived from re-running Model A with alternative specification 2 for conceptualizing and imputing the missing data. In this case, we find that 93% of the time the draws from the posterior are positive. While these results are not as robust as those in Models A and B, we claim that these results are not troublesome for our overall claims, particularly because our theory predicts a positive relationship making a one-tailed test more appropriate to analyze the within-agreement variable. Moreover, in this model we find the null results associated with across-agreement linkage found in Models A and B hold in Model C.

Table 1: Model A – Using Main Within-Agreement Linkage Specification

Within-Agreement Linkage:	3.15
	(1.91)
Across-Agreement Linkage:	
... in Related Policy	-.44
	(0.72)
... in Month	.15
	(0.67)
... in Council	.63
	(0.53)
Codecision	2.51
	(1.27)
Consensus Rule	5.03
	(2.04)
Regulation	3.53
	(1.46)
Constant	-1.63
	(1.50)

Coefficients are averaged across the five different Amelia datasets. Standard deviations are given parentheses. When we analyzed the posterior distributions we found that across the different models we found that only the within-agreement linkage measure, consensus rule, and regulation draws were consistently greater than zero more than 95% of the time. All three across-agreement linkage distributions consistently crossed zero.

Table 2: Model B – Using Alternative Within-Agreement Linkage Specification 1

Within-Agreement Linkage:	2.97
	(1.85)
Across-Agreement Linkage:	
... in Related Policy	-.44
	(0.70)
... in Month	.24
	(0.65)
... in Council	.57
	(0.52)
Codecision	2.60
	(1.29)
Consensus Rule	4.98
	(1.88)
Regulation	3.51
	(1.43)
Constant	-1.83
	(1.51)

Coefficients are averaged across the five different Amelia datasets. Standard deviations are given parentheses. When we analyzed the posterior distributions we found that only the within-agreement linkage measurement, consensus rule, codecision, and regulation draws were consistently greater than zero more than 95% of the time. All three across-agreement linkage distributions consistently crossed zero. This is only slightly less than in the full model presented in the paper.

Table 3: Model C – Using Alternative Within-Agreement Linkage Specification 2

Within-Agreement Linkage:	2.53
	(1.72)
Across-Agreement Linkage:	
... in Related Policy	-.43
	(0.70)
... in Month	.25
	(0.65)
... in Council	.55
	(0.52)
Codecision	2.60
	(1.28)
Consensus Rule	4.80
	(1.83)
Regulation	3.47
	(1.43)
Constant	-1.88
	(1.50)

Coefficients are averaged across the five different Amelia datasets. Standard deviations are given parentheses. When we analyzed the posterior distributions we found that only the within-agreement linkage measure, codecision, consensus, and regulation draws were consistently greater than zero more than 95% of the time. All three across-agreement linkage distributions consistently crossed zero. This is only slightly less than in the full model presented in the paper.

Going one step further to check the robustness of the results to the imputation of the missing data, **Model D** runs the analysis on a subset of the data – the proposals for which all data was available (N=34) (and thus no imputation was necessary). The results on this subset of complete data are reported in Table 4, and are consistent with the results reported in the paper. These results therefore provide additional support for our argument, showing that the effects associated with our key independent variable are not being driven by the imputed values of the missing data.

Table 4: Model D: Analysis of Full-Information Cases Only

Within-Agreement Linkage:	8.73
	(4.26)
Across-Agreement Linkage:	
... in Related Policy	1.03
	(1.06)
... in Month	-2.63
	(2.16)
... in Council	1.32
	(1.07)
Codecision	-3.40
	(3.60)
Consensus Rule	12.41
	(7.60)
Regulation	-2.55
	(2.44)
Constant	8.68
	(5.90)

Given the small n in the dataset with no missing data, we had to increase the burn-in period (10,000,000), amount of thinning (every 2000), and the length of MCMC chain (5,000,000) in order to ensure that there was no longer evidence of non-convergence.
Note that the coefficient on Consensus Rule is too large to be substantively meaningful.

Dealing with Possibility of Diffuse Reciprocity

To the control for the potential for diffuse reciprocity – linkages across agreements that take place over extended periods of time – we created three additional *Across-Agreement Council Linkage* measures. We expanded the time-frame of the original version of this measure measure, taking into account the potential for across-agreement linkages in the Council in which a particular agreement was adopted across a wider time period than simply the contemporaneous month. For example, for the $\pm$ one month measure, for a proposal adopted in October 2001 in agriculture we include a count of all agriculture proposals in September, October, and November 2001 as agreements for potential cross-agreement linkage. We created measurements to account for such possible diffuse linkages for one month, 1/4 year, and 1/2 year increments. The results of these diffuse reciprocity models are reported in Table 5. For all three diffuse reciprocity measures, we found that the posterior draws were too broad to state that there is a definitive relationship. Indeed, we find more support that the across-agreement council linkage measurement is actually more significant when we only include reciprocity within the month,

rather than in adjacent months. However, in these models, our within-agreement linkage measure remains positive more than 95% of the time for the one month and 1/4 year models, and positive more than 93% of the time for the 1/2-year model. (Note that in the 1/2-year model, this drop in significance is in part due to missing data on the across-agreement variable, which decreases the N, leading to greater uncertainty in the model.

Table 5: Controlling for Potential Diffuse Reciprocity Effects

	Model E	**Model F**	**Model G**	**Model H**
Within-Agreement Linkage:	3.15	3.08	3.37	3.63
	(1.39)	(1.85)	(1.87)	(2.48)
Across-Agreement Linkage:				
... in Related Policy	-.44	-.43	-.66	-3.56
	(.72)	(.71)	(.74)	(1.57)
... in Month	.15	.14	.19	1.06
	(.67)	(.67)	(.66)	(.94)
Diffuse Across-Agreement Linkage:				
... in (contemporaneous) Council	.63	—	—	—
	(.53)	—	—	—
... in Council, ± 1 month	—	.32	—	—
	—	(.43)	—	—
... in Council, ±1/4 year	—	—	2.89	—
	—	—	(.51)	—
... in Council, ±1/2 year	—	—	—	.47
	—	—	—	(.64)
Codecision	2.51	2.55	1.92	1.50
	(1.27)	(1.25)	(1.27)	(1.69)
Consensus Rule	5.03	4.89	4.71	7.34
	(2.04)	(1.86)	(1.84)	(3.03)
Regulation	3.53	3.31	2.89	7.51
	(1.46)	(1.43)	(1.44)	(3.19)
Constant	-1.63	-1.39	-.73	-1.40
	(1.50)	(1.51)	(1.89)	(2.61)

Coefficients are averaged across the five different Amelia datasets. Standard deviations are given parentheses.

Frequentist Results

Finally, we do not use the frequentist approach in the paper because our sample size is (too) small by frequentist standards. Indeed, Long (1997) recommends having an n of at least 100 in order to use a maximum likelihood estimator, because a small n can bias the results, resulting in the potential for type II errors. However, for the sake of transparency, we present here the results of frequentist logit models. Specifically, **Models I** through **L** re-run the original Models 1–4 from the paper using frequentist, rather than Bayesian analyses. The results of these models are reported in Table 6. The results show that there are very few differences in the results between these frequentist models and our Bayesian analyses. In the models focused on Cross-Agreement Linkage across Related Policies (Model J) and Cross-Agreement Linkage within a particular Month (Model K) – the key measures focused on policy and temporal linkage, in particular – we find that the effect of micro-linkage is positive, and statistically significant. In these cases, using the Bayesian model therefore allows us to infer that there is a positive relationship between within-agreement linkage and consensus, where the frequentist model would fail to do so, due to the small size of the dataset. In the model focused on Cross-Agreement Linkage within a given Council (Model L) and the model with all three measures (Model I), within-agreement linkage is not statistically significant at conventional levels. However, this is likely due to the fact that we are missing several values for this particular measure, meaning that the test is being run on even fewer than 70 cases – causing even more problems for inference in the frequentist approach (and consistent with the type II error we would expect to see in such small samples).

Table 6: Frequentist Logit Models

	Model I	**Model J**	**Model K**	**Model L**
Within-Agreement Linkage:	2.45	2.64*	2.65*	2.51
	(1.62)	(1.53)	(1.52)	(1.54)
Across-Agreement Linkage:				
... in Related Policy	-.32	-.21	—	—
	(.63)	(.57)	—	—
... in Month	.16	—	.03	—
	(.58)	—	(.49)	—
... in Council	.45	—	—	.24
	(.46)	—	—	(.40)
Codecision	1.89*	1.70*	1.58	1.65
	(1.07)	(1.01)	(.96)	(1.01)
Consensus Rule	3.65**	3.92**	3.94**	3.55**
	(1.5)	(1.44)	(1.33)	(1.41)
Regulation	2.63	2.67	2.56	2.49**
	(1.20)	(1.13)	(1.03)	(1.04)
Constant	-1.25	-.67	-.92	-1.07
	(1.30)	(0.96)	(1.14)	(.96)

* indicates p¡.10; ** indicates p¡.05.