Revision notes for "Intergroup Bias in Parliamentary Rule Enforcement"

March 16, 2016

Dear Editors of Political Research Quarterly,

Thank you for giving us the opportunity to revise our manuscript "Intergroup Bias in Parliamentary Rule Enforcement" (Manuscript ID PRQ-2015-0341). We greatly appreciate the time and effort taken by yourself and the four reviewers to review the manuscript. The comments and suggestions by yourself and the reviewers have been extremely helpful and we have made our best effort to accommodate the suggestions in the revised version of the manuscript. We believe the changes have made for a clearer, more complete, and more persuasive paper. We hope that you agree with us and deem the revisions satisfactory. However, we would, needless to say, be willing to make further revisions if desired.

Below we address the specific comments made by each of the reviewers in turn. These sections are labeled 1.1 for R1's first comment, 1.2 for R1's second comment, and so on. For ease of reading, we reproduce the reviewer comments before each of my responses.

In response to several of the reviewers' comments, we have added new figures and tables to the manuscript, some in the main text, some in the appendix. We reproduce these tables and figures below, along with other material only shown in this note.

Since the original submission of the manuscript, another debate has taken place. The analyses in the revised version of the manuscript includes data from the most recent debate, such that the number of remarks analyzed is 5,576 (up from 5,338 in the original version). The inclusion of the new data has not altered the results.

In accordance with PRQ's replication policy, we have also gathered reproduction materials (data and R scripts) in a GitHub repository. Footnote 4 in the revised manuscript provides a placeholder for the repository URL, though the exact URL is concealed in order to preserve anonymity.

Reviewer 1

1.1: Data to justify 'as if' randomness

First, the authors rely on the assumption that assignment of the chairperson can be considered "as if" random. It would be nice to have them provide some more empirical data so that we can evaluate that claim. At the very least I would like to see a graph that gave on the x-axis the percentage of seats in parliament controlled by the chairperson and on the y-axis the percentage of times they were chairing a co-partisan's speech. Ideally you should get something that looks like a 45 degree line.

We thank R1 for this suggestion. However, we do not believe the predicted pattern would constitute evidence in support of "as if" random assignment. The reason for this is that members of the leadership (with some exceptions for the head chairperson) are treated equally, with no consideration given to their party's share of seats in parliament. Hence, a chairperson from the 5th largest party has the same obligations as the chairperson from the largest party (again, with some exceptions for the head chairperson). Seat share should thus not be a predictor of chairperson activity. We want to stress that the manuscript was not sufficiently clear on this point, and we have added a paragraph in section 3.1 making this point more explicitly.

As additional evidence for the "as if" random assignment of chairpersons to debates, we have added section A.1, "Predicting chairman activity", to the appendix. Here we present results from regression models predicting how many debates chairpersons control. The key result from these models is that the main predictor of number of debates controlled is how long a chairperson has been a member of the leadership. There is an additional small premium for being the head chairperson, which likely reflects the fact that by convention, the head chairperson always takes the first shift controlling a debate. In contrast, number of seats has no significant association. The results are also shown below.

Table 1 shows that the main determinant of the number of remarks enforced by a chairmen is chairman's tenure, i.e. for how many debates he or she was in the leadership. In contrast, party seat share is uncorrelated with activity. The results complement the qualitative evidence from correspondence with the parliament leadership that chairperson activity is largely driven by 'supply-side' factors, i.e. chairpersons' availability on the day of the debate, and not by party-specific factors such as their parliamentary power.

1.2: Abrupt conclusion

Second, the conclusion is abrupt. While I like short papers, I think the authors can improve the impact of the paper if they can discuss more of the implications of the finding. What does this imply by how legislative institutions and rules are

Table 1: Model predicting number of remarks enforced by chairmen

	Number of remarks enforced		
	(1)	(2)	(3)
Head chairman	260.01**	235.37*	736.62***
	(121.61)	(126.67)	(249.60)
No. of debates in leadership	169.95***	167.01***	
	(19.52)	(20.06)	
No. of party seats		2.31	6.68
		(2.88)	(6.34)
Constant	−153.72**	−204.11**	14.28
	(54.74)	(83.63)	(177.87)
N	20	20	20
R^2	0.89	0.90	0.46
Adjusted R^2	0.88	0.88	0.39
Residual Std. Error	162.87 (df = 17)	164.60 (df = 16)	368.67 (df = 17)
F Statistic	71.52*** (df = 2; 17)	46.90*** (df = 3; 16)	7.12*** (df = 2; 17)

$^*p < .1$; $^{**}p < .05$; $^{***}p < .01$

formed? What are plausible things that might mitigate the bias? Does the answer to that question depend on the mechanism? If so, how? I think that exploring these questions, would be a way to increase the impact of the paper.

Based on this comment, we have extended the conclusion section (section 6) with more discussion of the findings's implications. For the specific wording of the discussion, we refer to the manuscript. In short, we argue that the findings imply that intergroup bias can occur even in institutional and cultural settings that should work strongly against bias (the same reasons for which we argue the setting is a 'least likely' case). At the same time, the observed effect is indeed fairly small, suggesting that the aforementioned factors do go some way towards mitigating bias. Lastly, and perhaps most importantly, they illustrate the importance of balancing institutions. As noted in section 5.2, the effects on speaking time for each chairperson roughly wash out within each debate due to the rotation of chairpersons. This does not nullify the finding per se, but illustrates how the principle of rotating chairperson roles can result in (approximate) equity in the aggregate even in the presence of bias at the individual level (see also the discussion in section 3.3 below).

We think this added discussion has strenghtened the manuscript, and we thank R1 for the suggestion.

1.3: Minor points

Small points: 1) In footnote 1, I think the authors intended to write "modal", not "model".

We have fixed the word in footnote 1 to 'modal', as was indeed intended.

2) Section 5.2 seems unnecessary.

We agree with this point in principle. However, section 5.2, which discusses the magnitude of the effect, speaks to a criticism sometimes raised against the finding: namely, that the observed effect is fairly small. This is indeed the case, and in section 5.2, we make an effort to be explicit about this point so as not to appear to overstate the magnitude of the effect. Hence, we have opted to keep section 5.2, though we agree that it could be omitted if need be.

Reviewer 2

2.1: Data to justify 'as if' randomness

[1] One thing that I think would be really important to show evidence of is the quasi-random assignment of the combination of speaker and chairman. My main concern was that there might be issues that are of particular importance to certain political parties. If this is true, then individuals from that party might try harder to be the chairmen when those issues are being debated. These might also be the same issues for which individuals from that party will naturally speak longer leading to a spurious own-group bias.

I believe that what you are finding is in-fact own-group but I also think the alternative explanation above is reasonable. It would be great to provide some more information about how some quantifiable characteristics of the issue or debate differ based on the party of the chairmen (this will show that chairmen appear to be randomly assigned across debates). You can also do this same comparison at the speech level and show that the characteristics of the debate are similar for the co-partisan and non-copartisan observations.

This is indeed a relevant concern, and we thank R2 for pointing it out. Because of the structure of the data, which only contains information about the duration of each remark and not its content, we are unfortunately unable to compare the issues covered by different speakers.

However, two aspects of how debates are organized indicate that issue-based self-selection is not likely. First of all, opening or closing debates are not about any one issue, but cover a multitude of issues, mostly based on issue disagreements between parties. MP's may debate one issue for a few remarks, and then shift to another issue. Second, the schedule of which chairmen oversee the debate is set in advance of the debate. Hence, chairmen are not able to opt in to the role when a particular issue comes up in the debate. The combination of these two aspects suggests self-selection based on issues is unlikely.

We have added a paragraph in section 3.1 of the manuscript making these points explicitly. For additional evidence in favor of the "as if" randomness assumption, see also section 1.1 in this note.

2.2: Debate-specific fixed effects

> *[2] One approach that might help allay this concern is to employ debate-specific fixed effects. You note on page 6 that the debate lasts for 12 to 16 hours and can include 600-800 brief remarks. Given the length of the debate, there are likely to be changes in the chairmen during the course of the debate. This would allow you to hold the issue at stake constant and see if the own-group bias persists net of this control.*

We thank R2 for this suggestion, which implies looking only at within-debate variation in who holds the chairmanship. We do want to note that a single debate typically covers several political issues, so adding debate-specific fixed effects does not strictly speaking imply holding issue content constant. Nevertheless, debate-specific fixed effects is a valuable robustness check of the results, which adds to the credibility of the result.

Based on this suggestion, we have added a table with the main results with debate-specific fixed effects added in every specification (Table 7 in section A.2 in the appendix). The results are robust to including debate-specific fixed effects. The table is also shown below. We mention this robustness test in the added section 5.4 "Additional tests".

Table 2: Results with debate-specific fixed effects

	Speaking time (seconds)				
	(1)	(2)	(3)	(4)	(5)
Copartisan	2.74***	2.63***	2.57***	1.42**	1.65***
	(0.88)	(0.90)	(0.93)	(0.55)	(0.53)
Time of day		−0.14	−0.13	−0.08	−0.08
		(0.13)	(0.13)	(0.13)	(0.14)
Gender (female)			0.53	0.04	0.11
			(0.51)	(0.51)	(0.58)
Intercept	58.53***	60.92***	60.58***	46.38***	47.65***
	(1.83)	(3.16)	(3.16)	(3.41)	(3.89)
N	5,756	5,756	5,756	5,756	5,756
Speaker party FE			✓		✓
Chair party FE					✓
Debate FE	✓	✓	✓	✓	✓
R^2	0.09	0.09	0.09	0.12	0.12
Adjusted R^2	0.09	0.09	0.09	0.12	0.12
chi^2	552.21*** (df = 10)	557.29*** (df = 11)	558.43*** (df = 12)	725.40*** (df = 21)	760.44*** (df = 28)

$^*p < .1$; $^{**}p < .05$; $^{***}p < .01$

2.3: More information about speaking assignment rules

> *To this point, it would also be helpful to provide a little more information about the rules about how the chairmen is to decide who gets to speak in what order. It*

seems that another way the chairman can allocate more speaking time to his/her own party is to just call on more people from that party. I was surprised that there wasn't really a discussion or test of this form of bias.

R2 asks for more information about the rules about how the chairman decides who gets to speak in what order. As R2 rightly notes, chairman may be able to allocate more time to his/her own party by just calling on more copartisans to speak, and the original version of the manuscript does not discuss this possibility.

We have added a discussion of this issue at the end of section 3.3. In short, chairmen are able to call on copartisans more than non-copartisans, but to a limited extent, since MP's sign up to speak using an electronic queue system. Chairmen can change the order of the queue, but do so only to a limited extent, and only (formally) with the purpose of giving party spokespersons preference. It is important to note that even if chairmen do give copartisans preference, this would not confound the estimate of interest, since we estimate bias in speaking time *conditional* on speaking. We thank R2 for raising this issue, which we had not originally thought of.

2.4: Slight bimodality in Figure 2

[3] The other thing that is really puzzling is in Figure 2 where the distribution for the non-copartisan speech durations has such an odd looking pattern. It looks like it has an initial peak right around 40 seconds and then another one just past the time limit. I think that you make some arguments for why it has this shape in the text but this is not what I would have expected if there was just some average bias occurring by judges. In that case, we should see a larger distribution just before the cutoff and a missing chunk of the distribution just after.

The slight bimodality in Figure 2 received some discussion in the original manuscript, but as R2 rightly points out, it is not quite the shape one would expect given just some average bias by chairmen.

To explore this issue further, we have changed the figure to present the distribution for non-copartisans separately for members of the chairman's own parliamentary bloc and for members of the other bloc. Figure 1 shows the revised figure.
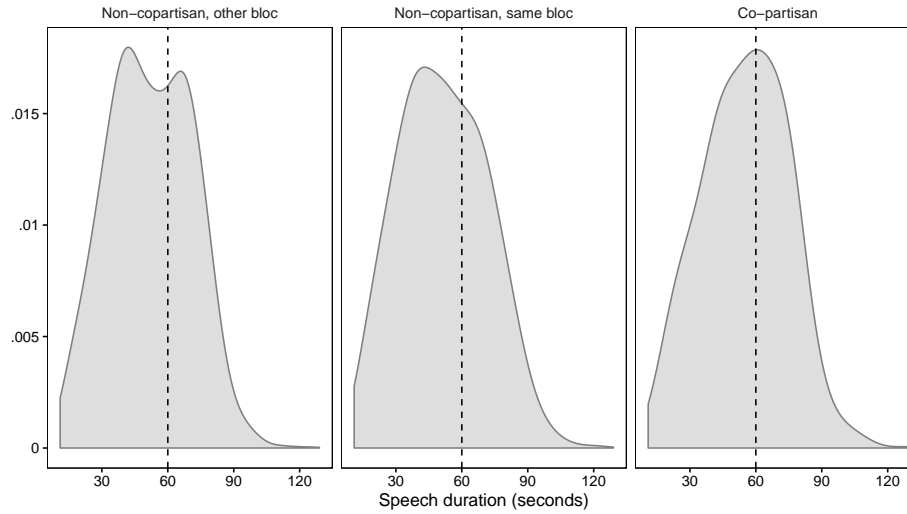
6

Figure 1: Distribution of speaking times for non-copartisans outside of the chairman's political bloc (left panel), non-copartisans from the chairman's own political bloc (middle panel) and copartisans (right panel).

Figure 1 makes clear that while the distribution of speaking times for same-bloc non-copartisans has the predicted shape, the shape for other-bloc non-copartisans is clearly slightly bimodal. In terms of explaining this slight bimodality, we lean towards the explanation proposed by R4 (see section 4.1 in this note for more discussion of this). In short, some non-copartisans may anticipate differential rule enforcement and thus constrain their remark, though we acknowledge that this leaves unanswered what accounts for this heterogeneity. This does in turn suggest that some of the observed difference reflects self-censorship. We have added a discussion of this to the beginning of section 5 of the paper.

Reviewer 3

We are aware that R3 is not fully convinced about the merits of the manuscript, and s/he raises a number of important concerns. As described below, we have made a number of changes to the manuscript in order to accomodate these concerns. As always, there are issues on which we do not entirely agree with R3, but we believe that we have good arguments for our dissent in these cases, and we discuss each below. We are not asking R3 to agree with us on these points, but we hope that s/he accepts our arguments. All in all we believe that the revised manuscript improves on the original in important ways. We hope R3 agrees with us on this and that we have succeeded in convincing her/him of the overall merits of the manuscript.

3.1: Analysis of distinction between rational choice and social identity processes

Unfortunately, the paper does not empirically show that the small co-partisan differences are due to social identity processes. There is some discussion about a rational choice perspective and identity processes, but there is no analysis shedding light on this distinction. But for some reason authors continue to attribute the reason to social identity.

We do not agree with the assessment that there is no analysis shedding light on the distinction between rational choice and social identity processes. Section 5.3 of the manuscript, testing the role of political moderators, has the purpose of exploring this distinction. As we argue, the fact that the bias is diminished when looking only within the chairman's own bloc is suggestive of a rational mechanism, though the test is not conclusive.

We appreciate R3's comment, and we have rewritten the introductory paragraph in section 5.3 to make this point clearer.

3.2: Partisanship as a 'soft' test

The paper is claiming that the results are the outcome of a natural experiment or that the results are obtained via a hard test of the impact of social identity in the Danish political context. Both are questionable. Partisanship of members of parliament is a very "soft" test for social identity, a "minimal group" would be a hard test. Parties for MPs are not minimal groups.

This point is well taken, and the test here is indeed softer than the test in Tajfel et al.'s original minimal group experiments. However, the contribution of this study is to test for intergroup bias in an ecologically valid setting. In contrast to experimentally assigned minimal groups, this study shows discrimination in naturally occurring groups. It is in the context of naturally occurring groups that we argue the setting constitutes a 'hard' test, insofar as it is characterized by transparency, limited political polarization and low levels of corruption. Hence, the study should be seen as a complement to minimal group experiments.

3.3: Pattern of queues of brief remarks

Regarding the claim about natural experiments, the fact that certain talks –most probably on not bi-partisan but polarizing issues- leads to a queue of brief remarks point to another impact not accounted for. Chairs are not randomly assigned to brief remarks, the nature of the initial talk generates the brief remarks.

R3 argues that "the fact that certain talks - most probably on not bi-partisan but polarizing issues - leads to a queue of brief remarks point to another impact not accounted for". This

reflects the very relevant concern that only some issues may generate brief remarks. In fact, every spokesperson speech generates a number of brief remarks, though this was not sufficiently clear in the original version of the manuscript. We thank R3 for highlighting this. We have rewritten section 3.1 of the manuscript, which describes the empirical setting, to clarify this.

The comment also touches on the issue of self-selection on the part of chairmen as well as MP's, which is indeed an important one. To better address concerns about self-selection, we have now expanded section 3.1 of the manuscript. Here we argue that while it is a valid concern, self-selection is unlikely to explain the observed results. In short, the schedule of presiding chairmen is set in advance of each debate, meaning that chairmen cannot self-select into shifts based on events during the debate. MP's sign up to give remarks based on the topic under discussion, and whether they were adressed in the remarks of another MP. By implication, they are effectively unable to time their remarks to coincide with a particular chairman.

3.4: Cancelling out of co-partisanship effect in the aggregate

> *Also, even the very small impact of co-partisanship is cancelled out because all chairs from different parties behave similarly. So, the article isn't even substantively interesting.*

R3 notes that the (small) impact of copartisanship cancels out in the aggregate because the chairmanship role rotates between parties. We thank R3 for highlighting this point, which was not made sufficiently clearly in the original version of the manuscript.

In the revised manuscript, we have added two paragraphs to the conclusion discussing this particular issue. While it is true that the biases will (roughly) tend to cancel out in the aggregate, we do not agree with R3 that this makes the finding substantively uninteresting. As we argue in the revised manuscript, it implies that "while institutional features such as simple rules and observability do not completely eradicate biased rule enforcement, a principle of rotation among rule enforcers can ensure that the impact of enforcement remains approximately balanced in the aggregate." We consider this an important implication, and it figures more prominently in the revised version of the manuscript.

3.5: Generalizability

> *This is at best a case study of the Danish Parliament. But the manuscript is claiming that results are generalizable because the perfectly functioning rule based internal workings of the Danish Parliament and the Danish political context somehow make the partisanship of MPs less salient. This is not what generalizability means.*

9

R3 takes issue with the claim of generalizability of the result. We appreciate this point, and we fully agree that the result is not generalizable in the sense of representing in any way a random sample of theoretically relevant settings. Instead, the result allows for 'least likely' case inference because intergroup bias is *ex ante* relatively unlikely in the setting studied here. In other words, the argument concerns theoretical rather than statistical inference.

We have rewritten the conclusion to make this distinction clearer. Specifically, the revised version of the conclusion does not use the word 'generalizability', which could be interpreted in the statistical sense. We thank R3 for pointing out this potentially misleading term.

3.6: Intentionality of social identity theory

> *p.4 social identity theory does not require or predict an "honest" wish to behave in line with rules. Speaking of external validity, I am not clear how the results obtained here in the Danish political context could be generalized to other parliaments.*

We thank R3 for pointing out this potentially misleading phrase. The purpose of the sentence in question was to highlight that social identity theory predicts intergroup bias even if *chairmen* may have an honest wish to treat speakers neutrally. It was not meant to imply that this is a specific theoretical requirement of social identity theory, but we see how it could be interpreted that way. We have rewritten the sentence in question to make this distinction clearer.

On the question of generalizability, see also the discussion of R3's earlier comment on this in section 3.5 in this note.

3.7: Part 5 vs. social identity theory

> *Part 5 is not compatible with the social identity perspective or cannot be explained by it.*

Part 5 of the manuscript is the entire results section, and we are not sure to which part of the results section R3 is referring in this comment. However, section 5.3 in the manuscript does explore the distinction between rational choice and social identity processes (see also section 3.1 in this note). Though the test is not conclusive, it is indeed the case that the evidence is somewhat suggestive of a rational (as opposed to a social identity) mechanism. In our view the manuscript as it is is reasonably clear on this point.

3.8: Intergroup bias vs. bias

> *Conclusion section says that the manuscript shows intergroup bias. It doesn't, it shows only bias.*

We are unsure of precisely the distinction referred to here, but we read this comment as interpreting "intergroup bias" to refer specifically to a social identity process. That is not the intended meaning. "Intergroup bias" does not specifically imply a social identity process, but is perfectly compatible with a rational process. In other words, a chairman who fully rationally discriminates against non-copartisan speakers (say, with the purpose of political gain) can also be said to show intergroup bias. It is our understanding that this represents the conventional use of the phrase "intergroup bias".

3.9: Reporting model significance

Please report model significance in Table 2 and elsewhere.

R3 requests that Table 2 and others report model significance. We omitted model significance statistics in the original version only for pragmatic, presentational reasons, since including them made the results tables too wide to fit on the page.

In the revised version, we have now squeezed the tables together a bit, and all the regression tables now report model significance (a χ^2 statistic). As shown, all model significance tests are strongly significant, i.e. the null of no improvement in fit compared to an intercept-only model can be rejected in all models.

3.10: Significance of 'prime minister' variable

Table 3- what does the significance of the "prime minister" variable mean?

R3 asks about the meaning of the significance of the "Prime Minister" (PM) variable. The question is very relevant, and we should have elaborated on the variable in the original version of the manuscript. In short, the PM variable captures remarks given when the MP speaking is also the sitting PM. The variable was included in the original manuscript in order to show that the effect was not driven by PMs, who may by virtue of their office be able to self-select into speaking slots (as opposed to other MP's, cf. section 2.3 in this note). As is clear from the original results, the PM variable is not a plausible confounder, as the copartisan variable remains significant throughout, and the PM variable is itself not consistently significant.

In the revised version of the manuscript, we have chosen not to include the PM variable. We have done so because upon further consideration we believe it is not theoretically relevant to distinguish between PM's and MP's in the context of this study: outside of their formal PM speeches (which are already excluded from the data), PM's have no special role or privileges in debates, and so they partake in debates only in their capacity of being MP's. For this reason, we have decided not to distinguish between PM's and other MP's in the revised version of the manuscript.

In any case, the PM's do not drive the results of the paper. (They are not particularly active during debates, and account for only about 6 percent of all remarks). In order to show this, we have added Table 8 in the appendix, which presents the results when PM remarks are excluded. As shown, the results are substantively unchanged. The table is also shown below.

Table 3: Results excluding prime ministers

| | Speaking time (seconds) | | | | |
	(1)	(2)	(3)	(4)	(5)
Copartisan	3.42***	3.14***	3.13***	2.11**	2.67***
	(1.00)	(1.16)	(1.12)	(1.00)	(0.72)
Time of day		−0.14	−0.12	−0.09	−0.06
		(0.14)	(0.15)	(0.14)	(0.14)
Gender (female)			1.39**	0.66	0.72
			(0.58)	(0.61)	(0.60)
Debate type (Opening)			−0.72	−0.38	0.34
			(1.08)	(0.93)	(0.92)
Intercept	51.45***	53.74***	53.21***	34.28***	36.66***
	(0.93)	(2.70)	(2.83)	(2.37)	(4.11)
N	5,379	5,379	5,379	5,379	5,379
Speaker party FE				✓	✓
Chair party FE					✓
R^2	0.004	0.005	0.01	0.04	0.06
Adjusted R^2	0.004	0.005	0.01	0.04	0.05
chi^2	22.50*** (df = 1)	26.62*** (df = 2)	35.22*** (df = 4)	222.21*** (df = 13)	308.80*** (df = 20)

$^*p < .1$; $^{**}p < .05$; $^{***}p < .01$

We thank R3 for directing our attention to this variable, which was insufficiently explained in the original version of the manuscript.

Reviewer 4

4.1: Interpretation

1. Interpretation. It seems to me that the authors provide no evidence that the presiding officers are gaveling non co-partisan speakers at the 60 second mark, but letting co-partisans (or ideologically proximate legislators) drone on for additional 2.8 seconds. So there are two possibilities here: the first is that MPs, for some reason, anticipate greater deference from copartisan presiding officers, and schedule *their* speeches to coincide with those presiding officers. Alternatively, perhaps those MPs want senior members of their own parties to hear particularly pressing points they have to make. The point is that while the timing of presiding officers may be effectively random, the scheduling of individual speeches need not be – so the match between MP making a speech and presiding

12

officer may have some selection bias. The author needs to convince the reader that this is not going on – that, for example, MPs get on the calendar before they know who is "up" for presiding officer, and can't reschedule.

The other possibility is that there is no selection bias, but that the MPs anticipate more favorable treatment from copartisan presiding officers once they are on the floor, and go on a little longer. Or, perhaps, copartisan MPs nod approvingly at the MPs talking points, and stare in stony silence at other MPs – and that this induces an unconscious bias on the part of the speaker which causes him/her to slow down or hurry up. I found the density plots in Figure 1 illuminating in this regard – note the mild bimodality among copartisans. It is entirely possible that body language on the part of the presiding officer might cause some MPs to slow down and others to speed up – but I would not say that this is equivalent to debates not being enforced equally – the authors' interpretation.

To sum up this comment, R4 notes (entirely correctly) that we do not provide direct evidence that chairmen gavel non-copartisans faster than co-partisans. Excluding this possibility leaves two possible mechanisms:

1. MP's self-select strategically into the debate schedule
2. Copartisan chairmen induce an unconscious bias on part of the speaker by means of subtle signals such as body language

R4 argues that the second interpretation "is [not] equivalent to debates being enforced unequally".

We thank R4 for these considerations. However, we disagree with R4's argument in a few specific respects. First of all, while it is correct that we do not provide direct evidence of differential formal enforcement by chairmen, we do not believe this is sufficient grounds for concluding that it does not take place. Though gaveling is not recorded in the transcripts, it does happen (as can be witnessed in video recordings of debates), and so should be considered an at least possible mechanism by which the observed bias in favor of copartisans arises.

Of the two alternative mechanisms proposed by R4, the first is that MP's self-select into speaking based on the partisanship of the presiding chairman. This is indeed an important concern, and the possibility of self-selection cannot be completely dismissed out of hand. In the original manuscript, the issue is partly covered by the concluding discussion of whether MP's anticipate biased rule enforcement.

In the revised manuscript, we have added a discussion of this issue to the description of the empirical setting in section 3.1. In short, since MP's sign up to speak 'on the fly' during debates, and the content of debates is largely dictated by the content of the spokespersons' speeches, the opportunity for an individual MP to strategically select when to give remarks based on either chairman partisanship or debate content is minimal.

13

The other mechanism proposed by R4 is that subtle signals from the chairman such as body language, unconsciously perceived by the speaker, give rise to the observed difference, possibly in a differential way which gives rise to the slightly bimodal distribution of speaking times in the left panel of Figure 1 above. As R4 notes, "body language on the part of the presiding officer might cause some MPs to slow down and other to speed up".

We agree entirely with R4 that this is a plausible mechanism. We only disagree with R4 in R4's assessment that this does not amount to unequal enforcement of the rules. Though qualitatively different from consciously biased enforcement (i.e., the first mechanism discussed by R4), it still at the end of the day amounts to unequal enforcement insofar as it affects speakers differently based on their copartisanship with the chairman. In other words, "unequal enforcement" (or "bias") is a behavioral characterization, not a motivational one, and as such can still meaningfully be said to exist even if not necessarily consciously perceived by either chairman or speaker.

We want to stress that we found R4's discussion of this illuminating, and in the revised version of the manuscript we have extended the discussion in section 6 of how the finding should be interpreted. Specifically, the rewritten section 6 includes a discussion of the interpretation of the effect with a particular focus on the distinction between consciously vs. unconsciously unequal enforcement suggested by R4.

4.2: Unexploited information about effect heterogeneity

> *2. Unexploited information. There are five major parties and the position of presiding chair rotates among them. Is the bias that the author documents present irrespective of which party the presiding chair comes from? Are the results being driven by a single party, or by a single chair/MP match? The coefficient on the copartisan variable in the main regression analysis is masking a lot of potential heterogeneity that might help the author pin down the story.*

We thank R4 for this suggestion, which we had not originally thought of. To test for effect heterogeneity, we have estimated random effects models with the same specification as model 5 in the main results, but allowing for the coefficient on the copartisan variable to vary by either chairman party or individual chairmen. The results are summarized in Figure 2 and Figure 3. In both figures, the parties are ordered from left (top) to right (bottom) based on voters' estimates of party positions.
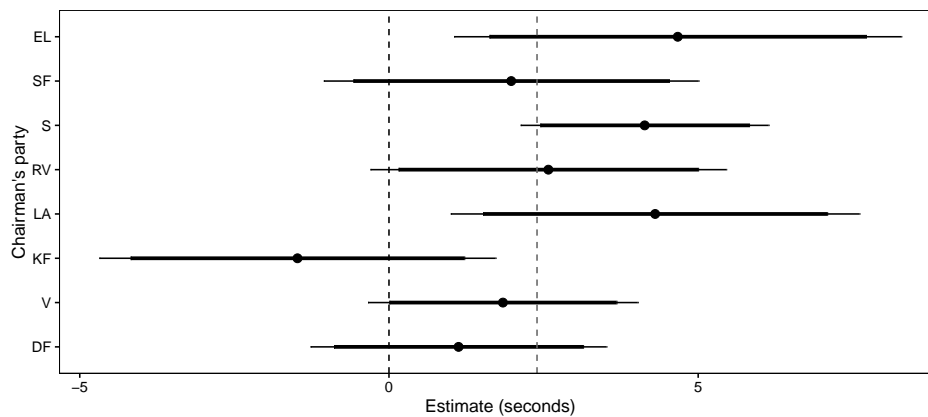
Figure 2: Estimates from a random effects model with varying slopes by chairman party. Effects are ordered by party position from left (top) to right (bottom).
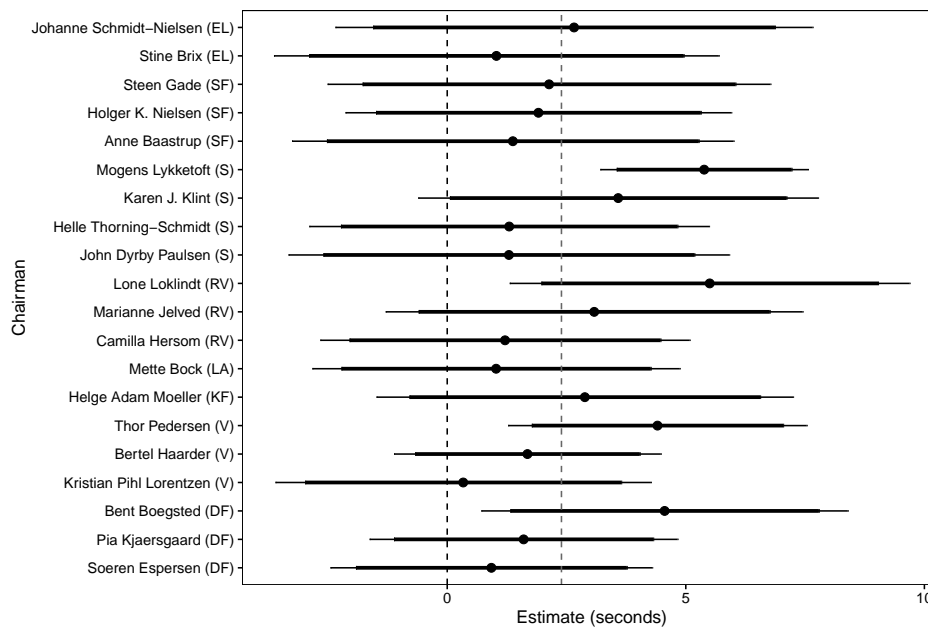

Figure 3: Estimates from a random effects model with varying slopes by chairman. Effects are ordered by party position from left (top) to right (bottom) and by descending coefficient value within parties.

Two lessons stand out in Figure 2 and Figure 3. First of all, looking at Figure 2, there appears to be no clear pattern with respect to effect heterogeneity across parties. If anything there may be a slight trend towards higher coefficients for more leftist parties, but the variance of the estimates is too high to conclude that with any certainty.

Second of all, though this trend is largely corroborated by Figure 3, there is one important

15

exception: one chairman, Mogens Lykketoft of the Social Democratic party (sixth from the top) is estimated to be significantly more biased in favor of copartisans than the average amount. This is significant because this study was originally inspired by allegations made by opposition parties in the summer of 2013 that the presiding head chairman was biased against the opposition (e.g., this article (in Danish)). The presiding head chairman was precisely Mogens Lykketoft. In our view the theoretically important point of the paper is the average bias, not the personal angle, but the particularly large estimate for Lykketoft does lend some additional face validity to the results.

For good measure, we have also tried to estimate model 5 in the main results on separate subsets of the data excluding each individual chairmen, to see if the results are driven by any single chairman. The estimates are shown in Figure 4
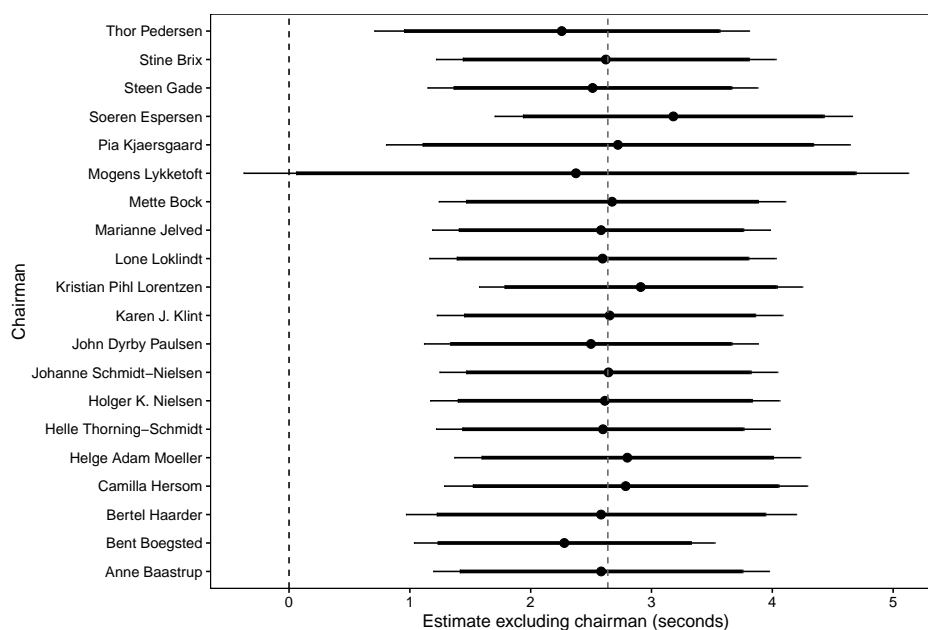

Figure 4: Estimates from model 5 in the main results on separate subsets of the data excluding each individual chairman. Effects are ordered by party position from left (top) to right (bottom) and by descending coefficient value within parties.

As shown in Figure 4, the results are robust to excluding each chairmen, demonstrating that the finding is not merely attributable to any single individual. The estimate for Mogens Lykketoft is less precise than the others because he alone accounts for nearly half of the observations, but the point estimate is in line with the others.

We have included Figure 2 and Figure 3 in the paper's appendix, and we describe the figures in the new section "Additional tests".