Capturing Rationalization Bias and Differential Item Functioning: A Unified Bayesian Scaling Approach

Jørgen Bølstad*

Abstract

Information about the ideological positions of different political actors is crucial in answering questions regarding political representation, polarization, and voting behavior. One way to obtain such information is to ask survey respondents to place actors on a common ideological scale, but, unfortunately, respondents typically display a set of biases when performing such placements. Key among these are rationalization bias and differential item functioning (DIF). While Aldrich-McKelvey (AM) scaling offers a useful solution to DIF, it ignores the issue of rationalization bias, and this study presents Monte Carlo simulations demonstrating that AM-type models thus can give inaccurate results. As a response to this challenge, this study develops an alternative Bayesian scaling approach, which simultaneously estimates DIF and rationalization bias, and therefore performs better when the latter bias is present.

Forthcoming in *Political Analysis*

Author's note: Replication materials are available on the *Political Analysis* Dataverse (Bølstad 2019). In order to facilitate future use, the replication folder includes a brief guide on how to use the provided functions. I have received many useful comments at various stages of this project, and I would like to thank Elin Haugsgjerd Allern, Pablo Barberá, Elias Dinas, Zoltán Fazekas, Bjørn Høyland, Carl Henrik Knutsen, Liam McGrath, Till Weber, Tore Wig, and the anonymous reviewers. I would also like to thank the Stan Development Team for their excellent support and advice. All remaining errors are my own.

*Associate Professor of Political Science, University of Stavanger, Norway. Email: jorgen.bolstad@uis.no.

1 Introduction

Without information about political actors' ideological positions, our ability to study politics would be severely limited. We would, for instance, not be able to assess whether the positions of political representatives reflect those of their voters, or examine the degree of polarization within each group. Neither would we be able to study the extent to which ideological proximity influences individual voting decisions, which has been a key question in electoral research. In other words, mapping the positions of different actors within a common ideological space forms a key step towards answering these and other important questions about politics.

In recent years, a number of approaches have been developed to perform such mapping. One approach relies on models for scaling legislators based on their roll-call votes (Clinton, Jackman, and Rivers 2004; Poole and Rosenthal 1985). By combining legislative voting data with survey responses from ordinary citizens regarding their positions on specific policy proposals, researchers have been able to estimate the positions of both legislators and citizens on the same scale (e.g. Bafumi and Herron 2010; Jessee 2009; Jessee 2016). However, a potential obstacle to using this approach is that it requires data from specifically tailored survey questions, which in practice may limit its feasibility. Another promising approach is to use social media data (e.g. Barberá 2015), but also this alternative has a few disadvantages. The sample will tend to be self-selected, and may still need to be matched with surveys to obtain data on other variables, such as vote choice.[1]

In this light, approaches relying only on electoral surveys are still attractive, as such surveys exist in virtually every democracy, and aim to cover a representative sample of voters. Electoral studies typically ask respondents to place themselves as well as political stimuli (such as parties or candidates) on the same ideological scale, which can be very useful. However, there are also well-known challenges to using such data. One of these is that respondents tend to interpret the scale differently, potentially disagreeing on what constitutes a centrist or moderate position, as well as how distances from this position should be mapped onto the scale. In other words, respondents shift and stretch the ideological space in different ways – a form of differential item functioning (DIF). In their classic contribution, Aldrich and McKelvey (1977) provide an approach to estimate and correct for DIF, and their approach is likely to see more use in the future, as new implementations have recently

[1] I do not aim to give an exhaustive literature review here, but simply highlight a few promising joint scaling approaches. Imai, Lo, and Olmsted (2016) discuss several models and provide fast expectation-maximization algorithms for situations with massive data.

become available (Poole et al. 2013; Hare et al. 2015).

While these implementations of the Aldrich-McKelvey (AM) model offer improved estimation procedures, the basic modeling assumption – the sufficiency of an intercept and a stretch parameter – has been retained for more than 40 years. This is potentially problematic because it ignores the pervasive effects of rationalization in the form of projection bias. That is, the tendency for respondents to place stimuli they like too close to themselves, and ones they do not like too far away (see, e.g. Johnston, Fournier, and Jenkins 2000). The failure to model these mechanisms is likely to bias the estimated shift and stretch parameters of the AM model, and hence also to bias the estimates of respondents' ideal points. In fact, the Monte Carlo simulations reported in this study show that the bias can be considerable. As one might expect, the stimuli position estimates are generally more robust to rationalization, but even these can be misleading.

This study addresses these issues by developing an alternative scaling approach, incorporating more information than the AM model to jointly estimate DIF and rationalization bias. As a by-product of this effort, the study also provides an improved version of the Bayesian AM implementation by Hare et al. (2015). The models are first tested and compared to existing models in Monte Carlo simulations. The differences between the models are then illustrated with an application to a difficult empirical case. Finally, posterior predictive checks demonstrate the degree of rationalization in the data and show the extent to which the models manage to fit such patterns.

2 Differential Item Functioning

A well-known challenge for inter-personal comparisons of survey data is that respondents tend to interpret questions differently (Brady 1985; King et al. 2004). Thus, respondents who are in perfect ideological agreement may place themselves at different positions on an ideological scale – or they may disagree considerably and nevertheless place themselves at the same position. In particular, it is likely that respondents disagree on two things: What constitutes the center of the scale (i.e. who the moderate actors are), as well as what a step on the scale means (i.e. how extreme the off-center actors are).

2.1 The Bayesian Aldrich-McKelvey (BAM) Model

Aldrich and McKelvey (1977) develop a model to capture such differences, estimating individual shift and stretch parameters that relate latent stimuli positions to reported perceptions of these. This approach has proven useful in a number of studies (e.g. Lo, Proksch, and Gschwend 2014; Saiegh 2009),[2] and it has also been tested and developed further.[3] Hare et al. (2015) develop a Bayesian implementation of the AM model (referred to as BAM), which allows for heteroskedasticity, and provides measures of uncertainty for all estimates of interest. The general question addressed in this literature is how respondent $i \in \{1, \ldots, N\}$ transforms the latent position θ_j of political actor $j \in \{1, \ldots, J\}$ into the reported position Y_{ij}. If we let $\phi(\cdot|\cdot, \cdot)$ be the probability density function of the normal distribution, then the likelihood function for the BAM model as specified by Hare et al. (2015) can be written as:

$$\prod_{i=1}^{N} \prod_{j=1}^{J} \phi(Y_{ij}|\alpha_i + \beta_i \theta_j, \sigma_{ij}^2), \tag{1}$$

where α_i is an intercept (or shift) parameter, β_i is a stretch (or weight) parameter, and σ_{ij}^2 is a variance term, constructed as $\sigma_{ij}^2 = (\tau_i \eta_j)^{-1}$ to allow for heteroskedasticity.[4] (Priors and further details on this model are discussed in a later section.) To calculate respondents' ideal points, the authors scale the reported self-placements (V_i) for each posterior draw as follows:[5]

$$v_i = \frac{V_i - \alpha_i}{\beta_i}. \tag{2}$$

Like other ideological scaling models, this approach assumes it is meaningful to map actors onto ideological dimensions that summarize preferences across a number of issue domains. This assumption underpins a large existing literature, using a variety of approaches (e.g. Bafumi and Herron 2010; Jessee 2009; Imai, Lo, and Olmsted 2016). However, a number of authors also question whether all voters hold strong ideological convictions (e.g. Barber and Pope 2019), and whether their issue preferences are sufficiently consistent for

[2] It is also an improvement over simpler approaches, such as replacing stimuli perceptions with mean placements, while otherwise ignoring DIF (e.g. Rabinowitz and Macdonald 1989).

[3] Notably, Palfrey and Poole (1987) questioned the assumption of homoskedastic errors, and simulated a scenario where this assumption was violated, concluding that the procedure was still performing well. Poole (1998) further introduced the blackbox method, generalizing the model to multiple dimensions and allowing missing data (see e.g. Bakker et al. 2014 for an application).

[4] The use of precision to specify the variance ($\sigma^2 = \tau^{-1}$) is consistent with the BUGS/JAGS language that the authors use. In this study, I translate the model to Stan, while retaining the authors' specification choices.

[5] To obtain point estimates of the voter positions, the authors to use the posterior median (because $\lim_{\beta_i \to 0} v_i = \infty$). The same approach is taken here. For stimuli positions, the posterior mean and median will typically give identical results.

summary measures to be meaningful (Broockman 2016). This concern is particularly relevant for studies of representation, which may be better off using issue-specific data (Ahler and Broockman 2018).

Nevertheless, ideological dimensions do play an important role in both political discourse and academic research, and the goal of this study is therefore to improve the models currently in use. The observation that some voters fit less well within spatial models may in fact also have relevant modeling implications. As discussed further below, the existing literature suggests that such voters still try to behave in accordance with a spatial model, by rationalizing their preferences. Such behavior would invalidate the assumption of the AM model that the shift and stretch parameters are sufficient to capture the differences in how respondents report their perceptions of political positions. In other words, such behavior would call for a more comprehensive model, which is what this study aims to develop.

3 Rationalization Bias

The notion of rationalization bias challenges the assumption that voters first form impressions of where political actors are located ideologically and then form preferences over these actors. It suggests a more complicated picture where respondents also adjust their (reported) perceptions of stimuli positions to fit their already formed political affinities and antipathies (Markus and Converse 1979; Conover and Feldman 1982; Krosnick 1990). The existing literature suggests that voters rationalize their preferences via two forms of projection (Merrill, Grofman, and Adams 2001; van der Brug 2001). First, respondents are likely to place stimuli that they like closer to themselves than they truly are, which is often referred to as *assimilation* effects. Second, respondents are likely to place stimuli that they do not like further away from themselves than they truly are, which is often referred to as *contrast* effects. Such behavior appears to be pervasive, posing a major challenge for studies using survey data on actor placements. In the words of Johnston, Fournier, and Jenkins (2000, 1151): "Perceptions are rife with bias".

Modeling such bias poses a couple of challenges. First, given the low number of stimuli in most settings, a parsimonious model is required: The number of individual-level parameters needs to be strictly limited. Second, the existing literature does not offer formalizations that can readily be included in a new model. In particular, contrast effects are somewhat complicated to model, and the existing literature contains very few attempts to do so.[6] I

[6] Merrill, Grofman, and Adams (2001), for instance, conduct a largely non-parametric analysis that

argue, however, that assimilation and contrast effects can be seen as facets of the same underlying phenomenon. As Westholm (1997) notes, what constitutes a good fit to individual preferences will depend on which spatial model respondents have in mind. Recent experiments suggest that most voters behave in accordance with a proximity model – appearing to penalizing distant actors and rewarding close ones (Tomz and Van Houweling 2008; Claassen 2007; Lacy and Paolino 2010). When voters' preferences diverge from this pattern, because they like actors that are ideological distant, or dislike actors that are close to them, cognitive dissonance theory (Festinger 1957; Aronson 1997) would predict that voters reduce this discrepancy by misperceiving the actors' stances as necessary.

I thus argue that the main consequence of both assimilation and contrasting processes is to exaggerate the fit of the proximity model of spatial voting. This point is not very clear in the literature on rationalization bias, but it can be very useful for modeling rationalization. While modeling assimilation and contrast effects separately might call for two new individual-level parameters, modeling these effects as part of the same underlying mechanism is more parsimonious. It also ensures that the effects are modeled in a theoretically and formally consistent way. The approach taken here is therefore to use a proximity model to define "ideal" stimuli positions and estimate the extent to which respondents move stimuli towards these positions.

3.1 The Intercept-Stretch-Rationalization (ISR) Model

The aim of this study is to build a model that accounts for rationalization bias as well as DIF. While the AM model is well suited for capturing DIF, accounting for rationalization bias requires a more elaborate model employing additional data. In particular, the model needs to incorporate information on whether the stimuli are liked or disliked by the respondents, as well as where the respondents are located.[7] Fortunately, surveys asking respondents to place themselves and political actors on ideological scales also tend to ask respondents for their

demonstrates assimilation and contrast effects without formalizing the processes. van der Brug (2001) offers a rare formalization, but not one that would be suitable for the present purposes.

[7] Like the existing AM-type models, the ISR model assumes that self-placements are not rationalized. While this cannot be tested with the data used here, a couple of points are worth noting. In the EES survey (which is analyzed below), respondents are asked to place themselves before the stimuli, which should make it less likely that they adjust the former placements to the latter. In general, the incentives to rationalize stimuli positions may also be stronger: Rationalization of self-placements is insufficient to achieve a perfect fit for the proximity model (except in very special cases), while rationalization of stimuli placements can generate a perfect fit for any set of preferences and any self-placement (without rationalization of the latter).

assessment of these actors (e.g. in terms of thermometer scales or probabilities of ever voting for them). In the following, individual i's assessment of stimulus j will be denoted U_{ij}, and the original data will be linearly transformed so that $U_{ij} \in [0, 1]$, where each respondents' least preferred stimulus is rated zero and the most preferred is rated one.[8]

As discussed above, I use the proximity model to define "ideal" stimuli positions, conditional on respondent self-placements and preferences, and estimate the extent to which respondents move stimuli towards these positions. As a first step, the standard proximity model can be defined as follows:[9]

$$U_{ij} = \kappa_i - \pi_i |p^*_{ij} - V_i|, \tag{3}$$

where κ_i is an intercept, π_i is a coefficient, and p^*_{ij} is a hypothetical party or candidate position. In the present case, $\kappa_i = \max(U) = 1$. The model can also be written somewhat differently, taking into account whether p^*_{ij} is on one or the other side of V_i, and letting π_i vary accordingly:

$$U_{ij} = \begin{cases} 1 + \pi_{1i}(p^*_{ij} - V_i), & \text{if } V_i > p^*_{ij} \\ 1 - \pi_{2i}(p^*_{ij} - V_i), & \text{otherwise.} \end{cases} \tag{4}$$

Assuming that preferences drop to zero when a stimulus reaches the upper bound (B) or the lower bound ($-B$), Equation 4 can be solved for π_{i1} and π_{i2}, which yields: $\pi_{i1} = (B+V_i)^{-1}$ and $\pi_{i2} = (B - V_i)^{-1}$.[10] Substituting these expressions for π_{i1} and π_{i2} in Equation 4 and solving for p^*_{ij}, we can define the "ideal" position for each stimuli as follows:

$$p^*_{ij} = \begin{cases} p_{1ij} = U_{ij}V_i + U_{ij}B - B, & \text{if } V_i > p^*_{ij} \\ p_{2ij} = U_{ij}V_i - U_{ij}B + B, & \text{otherwise.} \end{cases} \tag{5}$$

However, Equation 5 introduces a challenge, as it requires information on whether the

[8] An alternative is to simply set the maximum of the preference scale to one and the minimum to zero. However, this would limit the model's ability to detect high levels of rationalization for respondents who do not use the full scale.

[9] An alternative to this formulation is to use squared Euclidean distances: $(p^*_{ij} - V_i)^2$. However, such a model tends to fit observed data less well than the standard version with absolute distances (see, e.g. Merrill 1995, 283, Lewis and King 1999, 24, fn. 5).

[10] This is a slight deviation from the standard proximity model in which $\pi_{i1} = \pi_{i2}$. The reason for this deviation is both theoretical and practical: For all $V_i \neq 0$, if we assume preferences to reach zero at the most distant bound, and $\pi_{i1} = \pi_{i2}$, then they will only reach zero as we move beyond the bound on the other side. The current model yields predictions within the bounds for all scenarios.

respondent perceives the stimuli to be on one side or the other. We cannot, for instance, use the observed condition $V_i > Y_{ij}$ to define p^*_{ij} as this would bias the estimates of rationalization. Instead, we need an estimate of the direction rationalization will take. We could potentially use the condition $V_i > \theta^*_{ij}$, where θ^*_{ij} represents stimulus j's latent position translated to the respondent's own scale: $\theta^*_{ij} = \alpha_i + \beta_i\theta_j$. Yet this implies an abrupt shift as θ^*_j passes V_i, especially for the most disliked stimuli. This would in turn cause trouble for Hamiltonian Monte Carlo approaches, which require continuously differentiable functions. The alternative used here is inspired by the standard approach to handling discrete parameters when direct sampling is infeasible, namely marginalization. In particular, the model is specified as a mixture of two components, representing the two sides on which respondents may perceive the stimuli to be.

The model assumes that the direction of rationalization is at least partly determined by a latent DIF-process.[11] More specifically, it estimates the extent to which the mixing proportion $l_{ij} \in [0, 1]$ depends on a logistic transformation of the distance between V_i and θ^*_{ij}:

$$l_{ij} = \frac{\delta}{1 + e^{-z(V_i - \theta^*_{ij})}} + \frac{1-\delta}{2}, \tag{6}$$

where the parameter $\delta \in [0, 1]$ weights the DIF-based estimate against the alternative of equiprobability, and the constant z determines how rapidly the indicator shifts from zero to one.[12] The result is a smooth transition in the mixing proportions as θ^*_{ij} passes V_i.

To incorporate rationalization effects in the scaling model, the individual-level intercept and stretch parameters of the BAM model are supplemented with a rationalization component, resulting in a model that will be referred to as ISR. If we again let $\phi(\cdot|\cdot,\cdot)$ be the probability density function of the normal distribution, then the likelihood function for the ISR model can be written as:

$$\begin{aligned}\prod_{i=1}^{N}\prod_{j=1}^{J}[l_{ij}\phi(Y_{ij}|(1-\gamma_i)(\alpha_i+\beta_i\theta_j)+\gamma_i p_{1ij},\sigma^2_{ij})\\ +(1-l_{ij})\phi(Y_{ij}|(1-\gamma_i)(\alpha_i+\beta_i\theta_j)+\gamma_i p_{2ij},\sigma^2_{ij})],\end{aligned} \tag{7}$$

where the parameter $\gamma_i \in [0, 1]$ captures the extent of rationalization for each respondent.

[11] One could choose to be completely agnostic about the mixing proportions, and simply estimate their expectations from the data. However, this would separate the process of rationalization from DIF and likely ignore relevant information.

[12] In the analyses presented here, $z = 5/B$, which generally works well and gives a quicker transition for narrower scales.

4 Priors and Identification

It should be noted that the likelihoods discussed above are unidentified. Both the BAM model and the ISR model are characterized by reflection invariance and scaling invariance (see, e.g. Jackman 2001; Bafumi et al. 2005). Hare et al.'s solution is to fix two stimuli positions on the latent scale while placing wide, uniform priors on α and β: $\alpha_i, \beta_i \sim \text{Unif}(-100, 100)$. The authors further place standard normal priors on the remaining latent stimuli positions, $\theta_j \sim \text{Normal}(0, 1)$, and gamma priors on the precision terms: $\eta_j \sim \text{Gamma}(0.1, 0.1)$, and $\tau_i \sim \text{Gamma}(\nu, \omega)$. For the individual-specific precision terms, they also place gamma hyperpriors on the shape and rate parameters: $\nu, \omega \sim \text{Gamma}(0.1, 0.1)$.

However, Hare et al.'s approach would not be sufficient ensure identification of the ISR model. The approach taken here is rather to use more informative priors on α and β. Within the Bayesian paradigm, including prior information about the parameters is not only straightforward, but may also facilitate estimation and yield more accurate estimates (as the Monte Carlo simulations below also illustrate). The goal here is to include enough prior information to rule out unreasonable results and permit efficient estimation, while still allowing the results to be driven by the data (for similar arguments, see e.g. Gelman et al. 2008; Stan Development Team 2019).

In particular, β is given a normal prior with a mean of one, which implies an approximate one-to-one relationship between the observed and the latent scale. This is harmless because the latent stimuli positions are given a weakly informative prior, $\theta_j \sim \text{Normal}(0, 10^2)$, allowing for adjustment in accordance with the β's. A more important issue is the variance of the β-prior, which for the ISR model is set to one to allow for a considerable degree of stretching, while ruling out unreasonable values: $\beta_i \sim \text{Normal}(1, 1)$.[13] The intercept α is given a normal prior with a mean of zero, reflecting the assumption that both scales are centered at this point: $\alpha_i \sim \text{Normal}(0, \lambda^2)$. To permit a reasonable degree of shrinking, the standard deviation of this prior is given a weak half-Cauchy hyperprior: $\lambda \sim \text{Cauchy}^+(0, B)$.

Turning to the other parameters of the ISR model, δ is given a beta prior, reflect-

[13] When they first introduced their model, Aldrich and McKelvey (1977) noted that it perhaps would be most reasonable to constrain the stretch parameters to be positive, although the mathematical challenges this would entail made them reluctant to do so. Another reason was that they could not rule out that some respondents truly had what they referred to as negative weights, effectively perceiving the political landscape as a mirror-image where left is right and right is left. What their discussion illustrates, however, is that we can indeed express reasonable expectations regarding the parameter values. The prior used here takes into account that genuinely negative weights are far less probable than positive ones, as noted by Aldrich and McKelvey.

ing the belief that the direction of rationalization most likely depends on a DIF-process: $\delta \sim \text{Beta}(3, 1.1)$. The γ's are also given beta priors, with hyperparameters reflecting the estimated population distribution: $\gamma_i \sim \text{Beta}(\alpha_\gamma, \beta_\gamma)$, where $\alpha_\gamma, \beta_\gamma \sim \text{Gamma}(1.5, .5)$.[14] Finally, σ_{ij} is decomposed into the product of a unit J-simplex vector, ρ, and the parameter η_i, which represents the average variance of respondent i, implicitly increased by a factor of J^2.[15] In other words: $\sigma_{ij} = \rho_j \sqrt{\eta_i}$, where η_i is modeled with a scaled inverse chi-square distribution, $\eta_i \sim \text{Scale-inv-}\chi^2(\nu_\eta, \tau_\eta^2)$, using weakly informative hyperpriors: $\nu_\eta \sim \text{Cauchy}^+(0, 50)$, and $\tau_\eta \sim \text{Cauchy}^+(0, JB)$. The simplex vector is given a symmetric Dirichlet prior: $\rho \sim \text{Dirichlet}(5)$.

This means that the ISR model differs from the BAM specification by Hare et al. in two ways: (1) the inclusion of preference data to model rationalization, and (2) the choice of priors and the approach to identify the model. Because the model specified by Hare et al. is the most recent contribution to this literature, it is useful to test it as it is, and the Stan implementation of this model will simply be referred to as BAM. However, it also useful to examine whether the BAM model would perform better with different priors. The analyses below will therefore also include a version of the BAM model using the same priors as the ISR model, and this version will be referred to as BAM2. In addition, the analyses will include the original AM model, as this is useful for replicating existing studies (in particular Lo, Proksch, and Gschwend 2014). For convenience, the AM, BAM, and BAM2 models will be referred to as AM-type models.

5 Simulation Study

A key question is how well the AM-type models perform in the presence of rationalization. We would expect the estimates of voter positions to deteriorate as the degree of rationalization increases, and the same might apply to the estimates of stimuli positions. Another question is how the performance of the ISR model compares to the other models – whether rationalization is present or not. To answer these questions, this section reports a simulation

[14] The priors on α_γ, and β_γ, are also given a lower limit of one to ensure unimodality.

[15] Compared to multiplying ρ with J (and thus reducing the scale of η), this achieves faster sampling. The approach of decomposing the variance using a simplex vector is inspired by the R package rstanarm, but it is not the same as the one used there. The error structure is more complicated in the present case, and the approach outlined here tends to work well for the kind of two-dimensional, high-N-low-J heteroskedasticity we would typically expect to encounter. It is notably faster than the specification of Hare et al., where the individual components of σ_{ij}^2 are unidentified.

study.

The data in this study are generated using an ISR-type data generating process. I focus on three scenarios, where the γ-parameters are drawn from beta distributions producing increasing degrees of rationalization: $\text{Beta}(0, 1)$, $\text{Beta}(1, 3)$, and $\text{Beta}(1, 1)$. In other words, the first scenario entails no rationalization (only DIF), while the second and third scenarios entail γ-averages around .25, and .5, respectively. The results reported here entail moderately heteroskedastic errors with an average standard deviation close to .5, while a study with larger errors is reported as supplementary material. In both studies, the values of all other parameters are set approximately equal to those estimated in the empirical application below.

For each of the three scenarios, I generate 200 datasets of a typical size ($N = 500$, $J = 8$, $B = 5$). I fit the AM model using the maximum likelihood (ML) implementation in the R package basicspace. The BAM, BAM2, and ISR models are fit using a version of the No-U-Turn Sampler (NUTS) by Hoffman and Gelman (2014), which is an automatically tuned form of Hamiltonian Monte Carlo, implemented in Stan (Stan Development Team 2018, 2019; Carpenter et al. 2017). Because the models produce differently scaled results, I assess the voter and stimuli position estimates by their correlations with the true values.[16] Additional test criteria are reported as supplementary material.[17]

Figure 1 reports the main results. The left panel shows correlations between estimated and true stimuli positions. As we would expect, all models perform well in the first scenario, where there is no rationalization. Increasing the degree of rationalization moderately does not have a large impact on the performance of the AM-type models, although the tail gradually increases, reflecting poorer performance in a few cases. When we reach the scenario with the most rationalization, the performance of the AM-type models is notably worse than that of the ISR model, but it is still reasonably good in most cases.

The right panel shows similar correlations for voter positions. These estimates are particularly important to test, as they are hard to validate outside of a simulation context and are likely to be affected by rationalization. As we would expect, the AM-type models' ability to recover voter positions deteriorates with increasing rationalization, while the ISR model's performance is more stable. The correlations for the AM-type models are increasingly variable, with some results being very poor – to the extent that there is hardly any positive correlation at all. It is thus worth noting that even when the AM-type models perform well

[16] While I report Pearson's r, the results are substantively identical when the assumption of linearity is relaxed.

[17] Replication materials for the simulations and all other reported analyses are available on the *Political Analysis* Dataverse (Bølstad 2019).

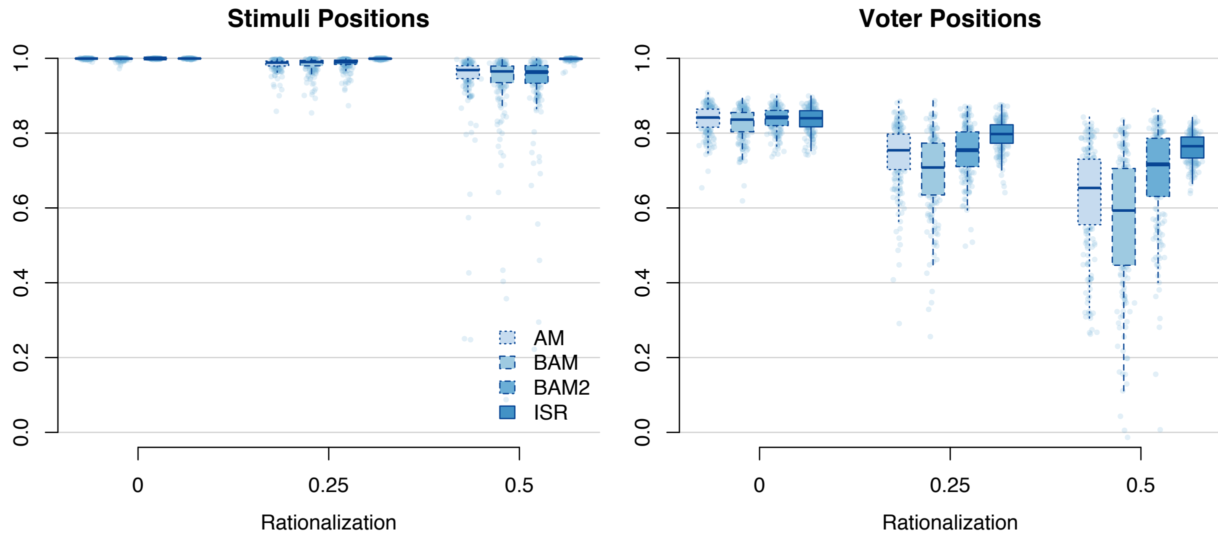

Figure 1. Correlations between estimates and true positions in simulations.

in terms of recovering stimuli positions (for which estimates are generally easier to validate), the voter position estimates may still be poor. Another point to notice is that the BAM2 model performs better than the two other AM-type models on this criteria, illustrating that more informative priors can make the model more robust to rationalization. The results from a similar simulation study with larger errors (reported as supplementary material) show that the BAM2 model is also more robust to random noise.

6 Empirical Application

Having seen how the models perform in a controlled setting, this section applies the models to real data. Among the existing studies employing AM-type models, Lo, Proksch, and Gschwend (2014) use a particularly comprehensive and suitable set of survey data, and these will be used here as well. The data come from the European Election Study 2009 (EES 2011), where the respondents were asked to place parties as well as themselves on 11-point left-right scales (see the supplementary materials for exact question wording).[18] To facilitate analysis and presentation, the scales are given a range from -5 to $+5$, with 0 as a center category. The EES respondents were also asked to report their party preferences,

[18] The odd number of categories implies a middle category, which has been found to increase the validity of such scales (Kroh 2007).

expressed as the probability of ever voting for a party.[19] In addition to the EES data, party position data from the Chapel Hill Expert Survey (CHES; Bakker et al. 2015) will be used for validation purposes.

Among the EES countries, I focus on a difficult case where the differences between the models are likely to be notable.[20] I examine the UK, which is a special case in several ways: The sample mean stimuli placements only have a correlation of .76 with the mean CHES placements, while most countries have correlations above .9. Furthermore, of all the countries Lo, Proksch, and Gschwend (2014) analyze, the UK is where their model is reported to explain the least of the variance – with the exception of Romania. In addition, 44% of the respondents are reported to have negative weights (again, only Romania with 46% has a higher share). This number is extreme, as it would suggest that nearly half the sample cannot tell the political left from right. In short, the AM model does not seem to fit very well, but Lo, Proksch, and Gschwend (2014) nevertheless obtain party positions estimates that are much more in line with expert placements than the mean placements are.

To ensure comparability, this case study uses the exact same data selection criteria as Lo, Proksch, and Gschwend, which yield a selection of 536 individuals.[21] Their results are replicated using the ML implementation of the AM model in the R package basicspace. However, in order to obtain outputs that are comparable to the other models (e.g. a posterior predictive distribution), the AM model is also implemented in a Bayesian fashion, using a model referred to as AM*. A key difference between the AM and BAM models is the AM model's assumption of homoskedastic errors, and the AM* model shares this assumption. Apart from this feature, the AM* model has the same specification and priors as the BAM model.[22]

6.1 Results

Table 1 summarizes the results. The ML-estimated AM model yields an exact replication of the results reported by Lo, Proksch, and Gschwend (2014), including 44% negative weights. The correlation between the party position estimates and CHES estimates is .92, which

[19] This is often referred to as a Propensity-to-Vote (PTV) question (Eijk et al. 2006; Eijk and Franklin 1996).

[20] Results for 14 other countries are provided as supplementary material.

[21] Due to their estimation procedure, Lo, Proksch, and Gschwend (2014) permit no missing values, and they require variance in each individual's party placements.

[22] It should be noted that the original AM specification is unsuited for Bayesian estimation. The BAM model's specification therefore differs somewhat from the original, and this also gives a slight difference between the AM and AM* models.

Table 1. Summary and validation of model fits for the UK.

Model	N	J	r, CHES	r, Means	$\hat{\beta}_i < 0$	WAIC	Max. $\hat{R}$
AM	536	8	0.92	0.49	0.44	NA	NA
AM*	536	8	0.92	0.56	0.42	20649	1.01
BAM	536	8	0.56	0.97	0.15	18834	1.01
BAM2	536	8	0.60	0.98	0.10	18632	1.01
ISR	536	8	0.98	0.75	0.14	17063	1.01

is an improvement over the naive approach of using sample mean placements (r = .76). Accordingly, the party positions estimates also differ considerably from the mean placements – the correlation between the two is only .49. The AM* model largely replicates these results, although with minor deviations.

The BAM model yields very different results. Interestingly, the share of negative weights is reduced from 44% to a more plausible 15%.[23] However, the party position estimates deviate notably from the CHES estimates, showing a modest correlation of .56. The party position estimates are instead very similar to the sample mean placements (r = .97), which would have been fine if these had greater convergent validity with the expert placements. We also see that the BAM2 model (which has the same priors as the ISR model) shows a similar performance to the BAM model, but has a slightly stronger correlation with the CHES estimates (.60) and fewer negative weights (10%). In sum, the homoskedastic AM models appear to produce more valid party position estimates, at the cost of an extremely high share of negative weights, while the heteroskedastic BAM models show the opposite characteristics.

While none of the AM-type specifications shows a fully satisfactory performance in this case, the ISR model produces more convincing results. The party position estimates have a very strong correlation with the expert placements, and the share of negative weights is more plausible than for the homoskedastic AM models. The table also reports the Watanabe-Akaike information criterion (WAIC; Watanabe 2010), which implies that the ISR model fits the data significantly better than the other models even when the effective number of parameters is penalized.

Given their considerable differences, the party position estimates of the BAM and ISR models deserve further attention, and these are shown in Figure 2. Interestingly, the BAM model appears to have particular problems recovering the positions of parties on the right

[23] This is calculated as the share of respondents whose posterior median β is below zero.

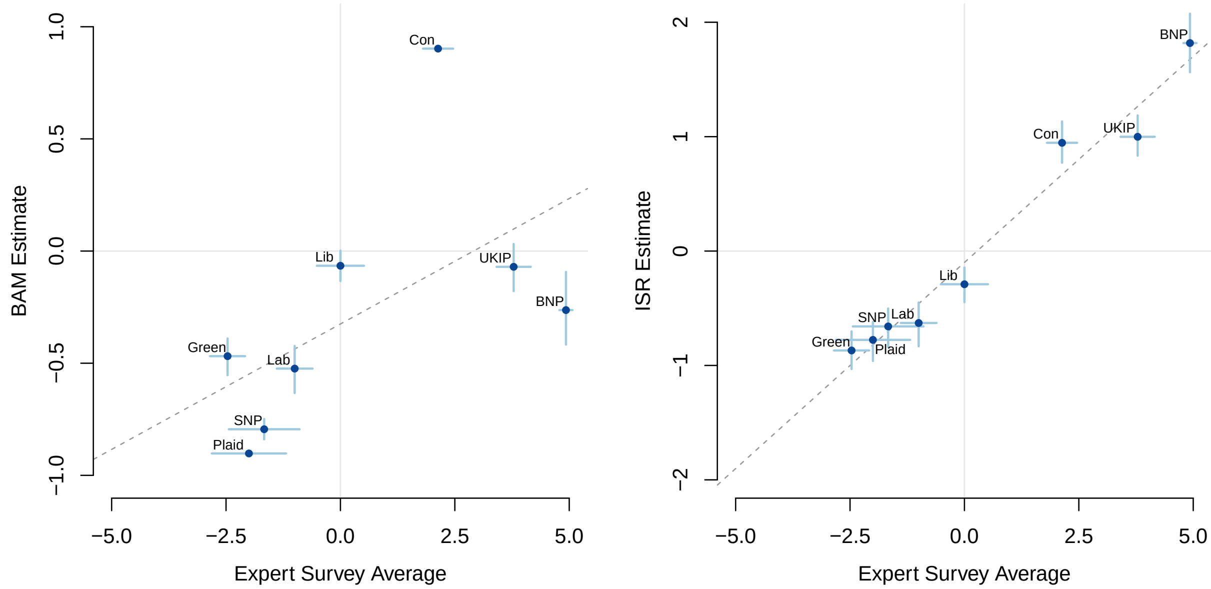

Figure 2. Party position estimates over expert placements.

side of the scale. The two parties that experts consider to be far-right – the UK Independence Party (UKIP) and the British National Party (BNP) – are estimated to be close to the center, while the Conservatives – which experts consider more moderate – are estimated to be more extreme. The BAM model also appears to place the Greens too close to the center, relative to the other parties.

Figure 3 further shows the relationships between the voter position estimates from three of the models, along with their densities. This illustrates the trade-off presented by the AM model, which in this case yields convincing party position estimates, at the cost of a large share of negative weights that are likely to give misleading voter position estimates in a number of cases. In fact, the voter estimates of the AM model only correlate with those of the ISR model and the BAM model at .42 and .38, while the correlation between the two latter is .67. (A similar plot including the BAM2 model is reported as supplementary material.)

A useful feature of the ISR model is that it allows us to assess the extent of rationalization in data, as well as whether there is heterogeneity in the tendency to rationalize. Figure 4 shows the relationship between the estimated γ-parameters and voter self-placements, by fitting penalized smoothing splines for each of 100 posterior samples. This reveals a pattern where voters on the left show only a weak to moderate tendency to rationalize, while those on the right do so to a much larger extent. This result is validated and discussed further

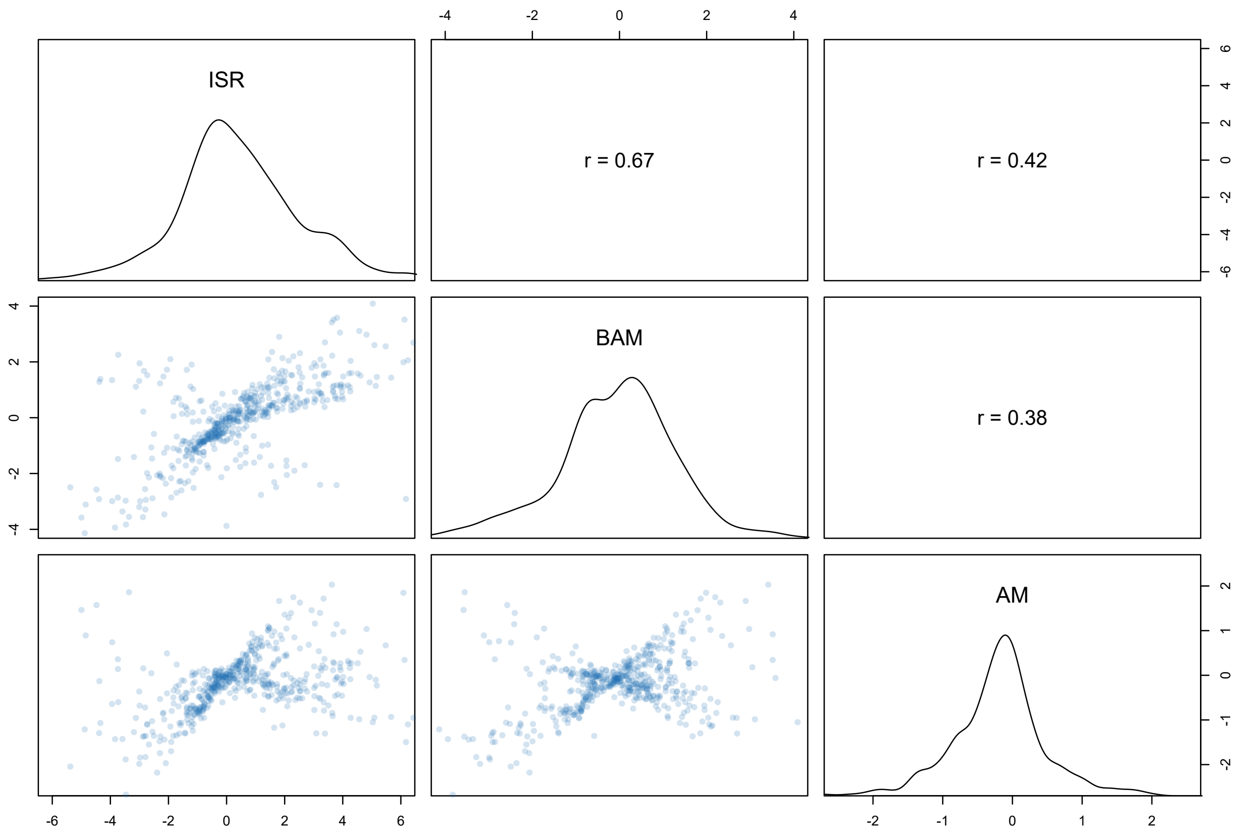

Figure 3. Correlations and densities of estimated voter positions.

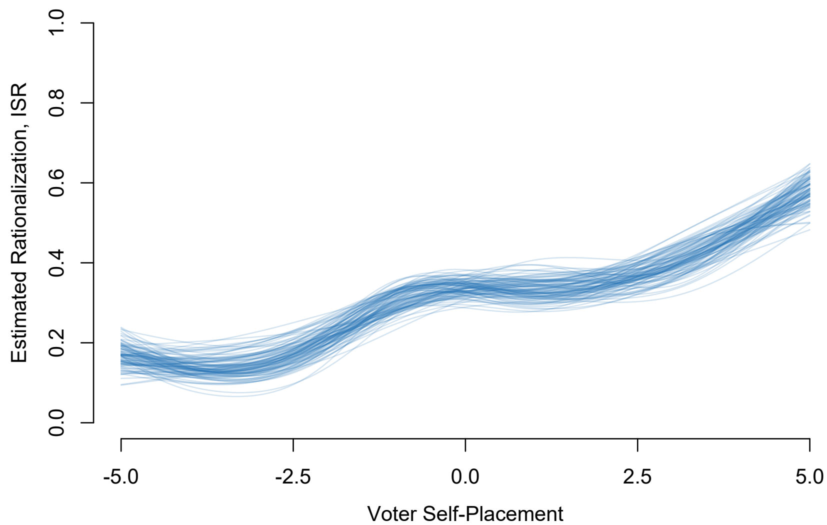

Figure 4. Estimated rationalization over voter positions.

below.

6.2 Observed and Predicted Patterns

This section checks how well the models fit the data, focusing on any observed patterns of rationalization. When a model fits well, data generated under the model should look similar to the observed data – this is the basic idea underlying posterior predictive model checking (Rubin 1984; Gelman et al. 2014). The tests below are based on 1500 draws of simulated outcome data from the posterior predictive distribution for each model. The key question is whether the predictions on average are consistent with observed response patterns across various combinations of self-placements, party positions and preferences.

In particular, Figure 5 reports the observed mean placement of each party along with densities of mean predictions, conditional on whether the respondents place themselves left ($V_i < 0$) or right of center ($V_i > 0$), as well as whether they like ($U_{ij} \geq .5$) or dislike ($U_{ij} < .5$) the party in question. With 8 parties, this gives 32 panels, which have been ordered so that the first row belongs to the leftmost party (according to CHES data), and the last row belongs to the rightmost. For reference, the average CHES placements are also shown with dashed vertical lines (these have been rescaled to range from -2 to 2, which is very close to the range of the observed EES survey means).

Figure 5 shows a remarkable degree of rationalization, but mostly among respondents on the right. As shown in the rightmost column of the plot, when respondents on the right

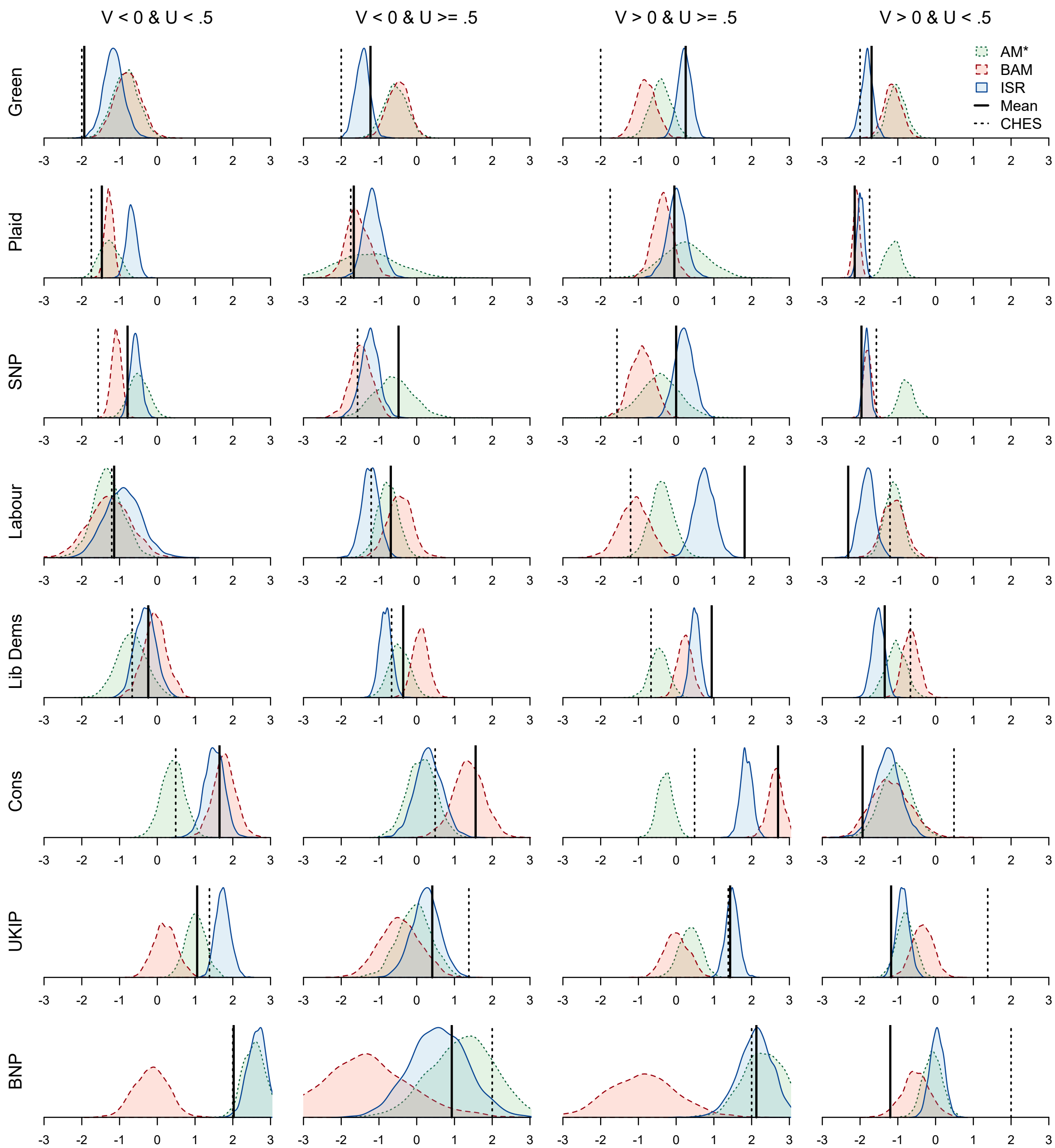

Figure 5. Posterior predictive densities and observed mean party placements.

dislike a party, they place it on the opposite side of the scale, no matter its actual position. Even the Conservatives, and parties such as BNP and UKIP – which experts consider to be far-right – are reported to be on the left by the rightist voters who do not like them. In other words, we see a clear example of contrasting. As shown in the third column of the plot, respondents on the right also show a notable degree of assimilation: When they like a party, they tend to place it on the right, even if experts place it firmly on the left. The key examples here are the Labour Party and the Greens (while SNP and Plaid Cymru are placed very close to the middle).

Among respondents who place themselves left of center (shown in the two leftmost columns), such patterns are much less pronounced. In fact, when it comes to placing parties on one or the other side of the center, these respondents agree with the experts in every single case. Comparing the two columns, we see that their placements depend considerably less on whether they like or dislike the parties – in most cases there is little difference. The clearest pattern of rationalization among these respondents is some degree of assimilation for BNP and UKIP, but overall they exhibit a much weaker tendency to rationalize than respondents on the right do. These observed patterns thus serve to validate the results shown in Figure 4, which also imply considerably more rationalization among respondents on the right than on the left.

The question is to what extent the models capture these patterns, and whether the observed data look plausible under each model. Ideally, we would like the observed means to lie within a range of high probability for a given model in every scenario. In practice, however, this would be a demanding requirement, as any uncaptured heterogeneity in how respondents treat specific parties may throw the predictions off. Indeed, what we see for the ISR model is that in those cases where its predictions are somewhat off, the observed mean tends to stands out compared to those of the neighboring parties. However, such cases are relatively rare for the ISR model – its predictions are in general close to the observed means. The two AM-type models show a less consistent performance, with very poor predictions in several cases. The predictions of the BAM2 model (which are reported as supplementary material) turn out slightly worse than those of the BAM model – which is not surprising for a more constrained model.

7 Conclusion

While survey data can provide very useful information on where voters and their representatives are located along ideological dimensions, we know that respondents often display a number of biases when answering surveys. The AM approach was developed to deal with the challenge of DIF, and it serves that purpose well. However, DIF is only one of the key biases survey respondents display, and this study aims to narrow the gap between our knowledge of political psychology and the models we employ.

The results presented above show that estimates of both voter and stimuli positions can be improved by modeling rationalization along with DIF. Monte Carlo simulations show that AM-type models produce increasingly inaccurate estimates of voter positions as the degree of rationalization increases. The effect on estimates of stimuli positions is considerably weaker, but can still be substantial – as the case of the UK also illustrates. In practice, when sample mean placements have high convergent validity (e.g. strong correlations with expert placements), the AM-type models may provide very accurate stimuli position estimates, but the ISR model still tends to fit the data significantly better.[24] Together with the Monte Carlo simulations, this implies that the voter position estimates from AM-type models are at risk of being biased even in cases where the stimuli positions are accurately estimated.

However, the results also show that the performance of AM-type models can be significantly improved by using more informative priors and an alternative approach to identify the parameters. The BAM2 model uses the same priors as the ISR model, and compared to the AM and BAM models, it is more robust to rationalization, as well as random noise. The BAM2 model should thus be preferred to the existing models, and could provide a relevant alternative to the ISR model – which requires more data. The ISR model is particularly useful when we are substantively interested in rationalization, or when we need estimates that are corrected for such bias. Yet, the ISR model performs as well as the BAM2 model even when there is no rationalization, so there is no clear reason not to consider it the default model.

It is, however, important to consider when these types of models are appropriate. A key assumption of the models examined here is that each stimulus has a single latent position on the scale, and this may seem uncontroversial if we agree that an actor can only have one position at a given time. The strength of these models is to map each actor within a common ideological space that exists beyond the actors' subjective perceptions. However, it is entirely possible that voters genuinely perceive party positions differently, because they

[24] See the results for 14 additional countries provided as supplementary material.

receive different messages, or interpret them in different ways (Lewis and King 1999). For some research questions, idiosyncratic perceptions (even biased ones) may provide the most appropriate data. As Bølstad and Dinas (2017) argue, this may be the case when we are more interested in subjective cognition than the role of external stimuli. In most other situations, however, ISR scaling should be a useful way to map voters and representatives on the same scale, based on a very common form of data.

In their conclusion, Hare et al. (2015, 772) note that "while the relationship between the Bayesian AM stimuli estimates and external measures of ideology [...] appears to be strong, it is not perfect", and ask "[w]hy are some stimuli perceived to be more or less ideologically extreme than is indicated by [such measures]? Are these deviations random noise or the result of ideological maneuvering?" This study may provide a piece of the answer: The deviations do not just represent random noise (or ideological maneuvering), but partly reflect rationalization bias. By taking this bias into account and providing more accurate estimates, the ISR model should facilitate further research on a number of topics in the areas of political representation and political behavior.

References

Ahler, D. J., and D. E. Broockman. 2018. "The Delegate Paradox: Why Polarized Politicians Can Represent Citizens Best". *The Journal of Politics* 80 (4): 1117–1133.

Aldrich, J. H., and R. D. McKelvey. 1977. "A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections". *The American Political Science Review* 71 (1): 111–130.

Aronson, E. 1997. "Back to the Future: Retrospective Review of Leon Festinger's 'A Theory of Cognitive Dissonance'". *The American Journal of Psychology* 110 (1): 127–137.

Bafumi, J., A. Gelman, D. K. Park, and N. Kaplan. 2005. "Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation". *Political Analysis* 13 (2): 171–187.

Bafumi, J., and M. C. Herron. 2010. "Leapfrog Representation and Extremism: A Study of American Voters and Their Members in Congress". *American Political Science Review* 104 (03): 519–542.

Bakker, R., C. De Vries, E. Edwards, L. Hooghe, S. Jolly, G. Marks, J. Polk, J. Rovny, M. Steenbergen, and M. A. Vachudova. 2015. "Measuring Party Positions in Europe: The Chapel Hill Expert Survey Trend File, 1999–2010". *Party Politics* 21 (1): 143–152.

Bakker, R., S. Jolly, J. Polk, and K. Poole. 2014. "The European Common Space: Extending the Use of Anchoring Vignettes". *The Journal of Politics* 76 (4): 1089–1101.

Barber, M., and J. C. Pope. 2019. "Does Party Trump Ideology? Disentangling Party and Ideology in America". *American Political Science Review* 113 (1): 38–54.

Barberá, P. 2015. "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data". *Political Analysis* 23 (1): 76–91.

Bølstad, J. 2019. "Replication Materials for: Capturing Rationalization Bias and Differential Item Functioning: A Unified Bayesian Scaling Approach". https://doi.org/10.7910/DVN/HXAKGJ, Harvard Dataverse, V1.

Bølstad, J., and E. Dinas. 2017. "A Categorization Theory of Spatial Voting: How the Center Divides the Political Space". *British Journal of Political Science* 47 (4): 829–850.

Brady, H. E. 1985. "The Perils of Survey Research: Inter-Personally Incomparable Responses". *Political Methodology* 11 (3/4): 269–291.

Broockman, D. E. 2016. "Approaches to Studying Policy Representation". *Legislative Studies Quarterly* 41 (1): 181–215.

Carpenter, B., A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. 2017. “Stan: A Probabilistic Programming Language”. *Journal of Statistical Software* 76 (1).

Claassen, R. L. 2007. “Ideology and Evaluation in an Experimental Setting”. *Political Research Quarterly* 60 (2): 263–273.

Clinton, J., S. Jackman, and D. Rivers. 2004. “The Statistical Analysis of Roll Call Data”. *American Political Science Review* 98 (2): 355–370.

Conover, P. J., and S. Feldman. 1982. “Projection and the Perception of Candidates’ Issue Positions”. *Western Political Quarterly* 35 (2): 228–244.

EES. 2011. “European Parliament Election Study 2009, Voter Study, Advance Release, 23/06/2011”. http://www.piredeu.eu.

Eijk, C. van der, W. van der Brug, M. Kroh, and M. Franklin. 2006. “Rethinking the Dependent Variable in Voting Behavior: On the Measurement and Analysis of Electoral Utilities”. *Electoral Studies* 25 (3): 424–447.

Eijk, C. van der, and M. N. Franklin. 1996. *Choosing Europe? The European Electorate and National Politics in the Face of Union*. Ann Arbor: The University of Michigan Press.

Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Evanston: Row, Peterson.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. *Bayesian Data Analysis*. 3rd edition. London: Chapman & Hall/CRC Press.

Gelman, A., A. Jakulin, M. G. Pittau, and Y.-S. Su. 2008. “A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models”. *The Annals of Applied Statistics* 2 (4): 1360–1383.

Hare, C., D. A. Armstrong, R. Bakker, R. Carroll, and K. T. Poole. 2015. “Using Bayesian Aldrich-McKelvey Scaling to Study Citizens’ Ideological Preferences and Perceptions”. *American Journal of Political Science* 59 (3): 759–774.

Hoffman, M. D., and A. Gelman. 2014. “The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo.” *Journal of Machine Learning Research* 15 (1): 1593–1623.

Imai, K., J. Lo, and J. Olmsted. 2016. “Fast Estimation of Ideal Points with Massive Data”. *American Political Science Review* 110 (4): 631–656.

Jackman, S. 2001. “Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference, and Model Checking”. *Political Analysis* 9 (3): 227–241.

Jessee, S. 2016. “(How) Can We Estimate the Ideology of Citizens and Political Elites on the Same Scale?” *American Journal of Political Science* 60 (4): 1108–1124.

Jessee, S. A. 2009. “Spatial Voting in the 2004 Presidential Election”. *American Political Science Review* 103 (1): 59–81.

Johnston, R., P. Fournier, and R. Jenkins. 2000. “Party Location and Party Support: Unpacking Competing Models”. *The Journal of Politics* 62 (4): 1145–1160.

King, G., C. J. Murray, J. A. Salomon, and A. Tandon. 2004. “Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research”. *American Political Science Review* 98 (1): 191–207.

Kroh, M. 2007. “Measuring Left–Right Political Orientation: The Choice of Response Format”. *Public Opinion Quarterly* 71 (2): 204–220.

Krosnick, J. A. 1990. “Americans’ Perceptions of Presidential Candidates: A Test of the Projection Hypothesis”. *Journal of Social Issues* 46 (2): 159–182.

Lacy, D., and P. Paolino. 2010. “Testing Proximity versus Directional Voting using Experiments”. *Electoral Studies* 29 (3): 460–471.

Lewis, J. B., and G. King. 1999. “No Evidence on Directional vs. Proximity Voting”. *Political analysis* 8 (1): 21–33.

Lo, J., S.-O. Proksch, and T. Gschwend. 2014. “A Common Left-Right Scale for Voters and Parties in Europe”. *Political Analysis* 22 (2): 205–223.

Markus, G. B., and P. E. Converse. 1979. “A Dynamic Simultaneous Equation Model of Electoral Choice”. *American Political Science Review* 73 (4): 1055–1070.

Merrill, S. I. 1995. “Discriminating Between the Directional and Proximity Spatial Models of Electoral Competition”. *Electoral Studies* 14 (3): 273–287.

Merrill, S., B. Grofman, and J. Adams. 2001. “Assimilation and Contrast Effects in Voter Projections of Party Locations: Evidence from Norway, France, and the USA”. *European Journal of Political Research* 40 (2): 199–221.

Palfrey, T. R., and K. T. Poole. 1987. “The Relationship Between Information, Ideology, and Voting Behavior”. *American Journal of Political Science* 31 (3): 511–530.

Poole, K. T. 1998. “Recovering a Basic Space from a Set of Issue Scales”. *American Journal of Political Science* 42 (3): 954–993.

Poole, K. T., and H. Rosenthal. 1985. “A Spatial Model for Legislative Roll Call Analysis”. *American Journal of Political Science* 29 (2): 357–384.

Poole, K., H. Rosenthal, J. Lewis, J. Lo, and R. Carroll. 2013. “basicspace: A Package to Recover a Basic Space from Issue Scales.” http://CRAN.R-project.org/package=basicspace.

Rabinowitz, G., and S. E. Macdonald. 1989. “A Directional Theory of Issue Voting”. *American Political Science Review* 83 (1): 93–121.

Rubin, D. B., et al. 1984. “Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician”. *The Annals of Statistics* 12 (4): 1151–1172.

Saiegh, S. M. 2009. “Recovering a Basic Space from Elite Surveys: Evidence from Latin America”. *Legislative Studies Quarterly* 34 (1): 117–145.

Stan Development Team. 2018. “RStan: The R Interface to Stan”. R package version 2.17.3. http://mc-stan.org.

Stan Development Team. 2019. “Stan User’s Guide”. Version 2.19. http://mc-stan.org.

Tomz, M., and R. P. Van Houweling. 2008. “Candidate Positioning and Voter Choice”. *American Political Science Review* 102 (3): 303–318.

van der Brug, W. 2001. “Perceptions, Opinions and Party Preferences in the Face of a Real World Event Chernobyl as a Natural Experiment in Political Psychology”. *Journal of Theoretical Politics* 13 (1): 53–80.

Watanabe, S. 2010. “Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory”. *Journal of Machine Learning Research* 11:3571–3594.

Westholm, A. 1997. “Distance Versus Direction: The Illusory Defeat of the Proximity Theory of Electoral Choice”. *American Political Science Review* 91 (4): 865–883.