This is an Accepted Manuscript of an article published by Taylor & Francis in Multivariate Behavioral Research on 29/5/2018, available online: http://www.tandfonline.com/10.1080/00273171.2018.1469966.

A New Model for Acquiescence at the Interface of Psychometrics and Cognitive Psychology

Hansjörg Plieninger and Daniel W. Heck
University of Mannheim

When measuring psychological traits, one has to consider that respondents often show content-unrelated response behavior in answering questionnaires. To disentangle the target trait and two such response styles, extreme responding and midpoint responding, Böckenholt (2012a) developed an item response model based on a latent processing tree structure. We propose a theoretically motivated extension of this model to also measure acquiescence, the tendency to agree with both regular and reversed items. Substantively, our approach builds on multinomial processing tree (MPT) models that are used in cognitive psychology to disentangle qualitatively distinct processes. Accordingly, the new model for response styles assumes a mixture distribution of affirmative responses, which are either determined by the underlying target trait or by acquiescence. In order to estimate the model parameters, we rely on Bayesian hierarchical estimation of MPT models. In simulations, we show that the model provides unbiased estimates of response styles and the target trait, and we compare the new model and Böckenholt's model in a recovery study. An empirical example from personality psychology is used for illustrative purposes.

Keywords: Acquiescence, Bayesian hierarchical modeling, item response theory (IRT), response styles, multinomial processing tree models

Introduction

Questionnaires with Likert-type response formats are widely used to assess various constructs such as personality variables, mental disorders, or attitudes towards products, teachers, or co-workers. Despite their widespread application, however, concerns have been raised about the validity of Likert-type data because of response styles. Such response styles are defined as systematic preferences of respondents for specific response categories that cannot be explained by the item content. Three prominent response styles are the "tendency to use positive response categories (acquiescence response style, or ARS), [...] the midpoint response category (midpoint response style, or MRS), and extreme response categories (extreme response style, or ERS)" (Weijters, Geuens, & Schillewaert, 2010b, p. 96).

Early psychometric models for response styles usually focused on a single response style, for example, mixture distribution Rasch models for ERS (e.g., Rost, Carstensen, & von Davier, 1997) or factor models for ARS (e.g., Billiet & McClendon, 2000). In recent years, models that account for more than one response style have been developed (e.g., Johnson & Bolt, 2010). For example, Böckenholt (2012a) proposed so-called item-response-tree (IR-tree) models to account for both MRS and ERS. This model has the appeal that it assumes a psychologically meaningful tree-like structure of the underlying processes. In the present manuscript, we extend this model to acquiescence by building on multinomial processing tree (MPT) models from cognitive psychology and recent computational advances in Bayesian hierarchical estimation of these models. In the remainder of the Introduction, we review the literature on ARS and develop the proposed model on the basis of IR-tree models and hierarchical MPT models. Next, two simulation studies address parameter and model recovery, respectively. Finally, an empirical example is used to illustrate the proposed approach.

Acquiescence

Interindividual differences in acquiescence have been a topic of research for almost a century (Cronbach, 1946; Lentz, 1938). ARS is usually defined in terms of observable response patterns, namely, systematically more *agree* responses compared to what would be expected on the basis of a person's target trait (e.g., Paulhus, 1991). Bentler, Jackson, and Messick (1971) referred to this as *agreement acquiescence*.[1] Research on ARS has often focused on correlates of acquiescence (see Van Vaerenbergh & Thomas, 2013; Wetzel, Böhnke, & Brown, 2016). For example, previous studies found relationships of ARS with age, socioeconomic status, cultural variables, impulsiveness, extraversion, or cognitive capacity (e.g., Austin, Deary, & Egan, 2006; Couch & Keniston, 1960; Johnson, Kulesa, Cho, & Shavitt, 2005; Meisenberg & Williams, 2008; Soto, John, Gosling, & Potter, 2008). However, as Wetzel et al. (2016) write, "due to inconsistencies in the findings of these studies—which may in part be attributed to differences in measuring response biases—there are not always clear results" (p. 349). Another stream of research has investigated the stability of ARS and found it to be relatively stable across content domains (e.g., Danner, Aichholzer, & Rammstedt, 2015; Weijters, Geuens, & Schillewaert, 2010a) and across periods of months or even years (e.g., Billiet & Davidov, 2008; Weijters et al., 2010b). However, there are also studies reporting a rather low consistency of ARS (e.g., Ferrando, Condon, & Chico, 2004), and Rorer (1965) even called acquiescence a "myth". This was rebutted by Bentler et al. (1971), and researchers today seem to agree that both content and, to a smaller degree, acquiescence influence questionnaire responses.

Apart from these empirical results, methodological approaches to measuring ARS differ substantially. The first, most prominent approach relies on the aggregation of responses across both regular and reversed items without recoding the latter. This idea is either used to form a manifest measure of ARS for partialling (e.g., ten Berge, 1999). Or, this idea is employed in a two-factor model that includes one factor for the target trait and a second factor for ARS with loadings of $+1$ and -1 for regular and reversed items, respectively (e.g., Billiet & McClendon, 2000; Maydeu-Olivares & Coffman, 2006; Mirowsky & Ross, 1991). It is important to note that this approach treats acquiescence and disacquiescence as opposite poles of a single dimension. A second approach to measuring ARS can be seen as a special case of the first approach. It is based on pairs of logical opposite items (e.g., "I am happy" and "I am sad") and ARS is simply the mean across such items (e.g., Soto et al., 2008; Winkler, Kanouse, & Ware, 1982). A third approach focuses only on the *agree* categories and was used with content-heterogeneous items (e.g., Weijters et al., 2010b) or with content-homogeneous items (e.g., Falk & Cai, 2016; Johnson & Bolt, 2010; Wetzel & Carstensen, 2017). Even though it is obvious that the approaches differ in their theoretical definition and operationalization of ARS, they are seldomly compared, and empirical evidence of convergent validity is mixed (Billiet & McClendon, 2000; Ferrando et al., 2004; Kam & Zhou, 2015).

Another theoretical question concerns the relationship between acquiescence and related phenomena such as item-wording effects or careless responding. Item-wording effects are usually modeled as a method or residual factor to account for shared variance among negatively worded items (e.g., Marsh, Scalas, & Nagengast, 2010). Thus, the approach has some overlap with the two-factor model for measuring ARS even though

Hansjörg Plieninger, School of Social Sciences, Department of Psychology, University of Mannheim, Germany; Daniel W. Heck, School of Social Sciences, Department of Psychology, University of Mannheim, Mannheim, Germany.

This work was supported by the University of Mannheim's Graduate School of Economic and Social Sciences funded by the German Research Foundation (DFG) and the research training group "Statistical Modeling in Psychology" (GRK-2277), also funded by the DFG.

This work is partly based on the first author's doctoral dissertation (Hansjörg Plieninger, 2018).

The authors would like to thank Morten Moshagen, Benjamin Hilbig, and Ingo Zettler for generously providing their data for secondary analysis as well as Thorsten Meiser for helpful comments on a previous version of this manuscript.

Correspondence concerning this article should be addressed to Hansjörg Plieninger, Department of Psychology, University of Mannheim, 68131 Mannheim, Germany. E-mail: plieninger@uni-mannheim.de

[1] According to Bentler et al. (1971), *agreement acquiescence* is the tendency to agree, whereas *acceptance acquiescence* is the tendency to accept items as self-descriptive. Persons high on the former are predicted to agree with all items, whereas persons high on the latter are predicted to agree with descriptions (e.g.,"happy", "sad") and to disagree with denials (e.g., "not happy", "not sad").

only few studies compared the two phenomena (e.g., Weijters & Baumgartner, 2012; Weijters, Baumgartner, & Schillewaert, 2013). Apart from that, respondents that are inattentive or careless and thus miss reversals or negations of items are sometimes described as careless responders. Even though careless responding and ARS may both inflate the number of *agree* responses, acquiescence is not necessarily related to being inattentive (Swain, Weathers, & Niedrich, 2008; Weijters & Baumgartner, 2012; Weijters et al., 2013).

From a cognitive perspective, the underlying process of responding to questionnaire items has been described by the four stages of interpretation, retrieval, judgment, and responding (Shulruf, Hattie, & Dixon, 2008; Tourangeau & Rasinski, 1988; Zaller & Feldman, 1992). Weijters et al. (2013) stated that acquiescence is related to category usage and is thus a problem associated with the last stage. In contrast, Knowles and Condon (1999) proposed that acquiescence is related to earlier stages, because it involves a failure to reconsider an initially accepted item. This latter position is in line with the concept of agreement acquiescence (Bentler et al., 1971) and the theory on satisficing, which states that acquiescence is the result of an impaired response process due to factors such as cognitive or motivational restrictions (Krosnick, 1999).

In summary, this short review—with a focus on correlates, stability, measurement, related phenomena, and cognitive response processes—shows that the literature on acquiescence is far from unanimous in these respects. In the present paper, we adopt the position that acquiescence is a consistent, trait-like construct that leads to systematically more *agree* responses. By developing a new model based on these assumptions, we aim to shed new light on the theoretical definition, measurement, and underlying processes of acquiescence. However, the proposed model cannot reconcile all of the diverse and often unresolved questions concerning the phenomenon of ARS. In the following sections, we will first develop a new model of acquiescence before elaborating on its theoretical implications.

The IR-Tree Model of Response Styles

Böckenholt (2012a), as well as De Boeck and Partchev (2012), developed the class of IR-tree models, thereby generalizing sequential item response models that have been proposed for polytomous items (Tutz, 1990; Verhelst, Glas, & de Vries, 1997). Herein, we will focus on a specific IR-tree model, namely, a response style model for questionnaire items with an ordinal, symmetric 5-point response format, henceforth called the *Böckenholt Model*. Usually, questionnaires with such items are analyzed using unidimensional, ordinal models. However, in the Böckenholt Model, it is assumed that three distinct processes account for the observed responses (see Figure 1): First, a person i ($i = 1, \ldots, I$) may enter an MRS stage on item j ($j = 1, \ldots, J$) with probability m_{ij} and thus give a midpoint response. Otherwise, the complementary stage is entered with probability $(1-m_{ij})$, in which case the content of the item is evaluated. The person is assumed to enter a latent state of agreeing or disagreeing with the item's content with probability t_{ij} and $(1 - t_{ij})$, respectively, depending on the person's target trait and the item's difficulty. For example, if the items are designed to measure happiness, the states may be interpreted in terms of happy versus unhappy. Finally, an ERS stage is entered with probability e_{ij} leading to a *strongly agree* response in case of agreement and a *strongly disagree* response in case of disagreement. The complementary stage is entered with probability of $(1 - e_{ij})$ leading to moderate *agree* and *disagree* responses, respectively.

The three model parameters m_{ij}, e_{ij}, and t_{ij} are parameterized using person parameters θ and item parameters β:

$$m_{ij} = \Phi(\theta_{mi} - \beta_{mj}) \tag{1}$$
$$e_{ij} = \Phi(\theta_{ei} - \beta_{ej}) \tag{2}$$
$$t_{ij} = \Phi(\theta_{ti} - \beta_{tj}). \tag{3}$$

Substantively, each person is assumed to have three latent traits $\boldsymbol{\theta}_i = (\theta_{mi}, \theta_{ei}, \theta_{ti})'$ and each item is modeled using three difficulty parameters $\boldsymbol{\beta}_j = (\beta_{mj}, \beta_{ej}, \beta_{tj})'$. Individual differences with respect to the target trait (e.g., happiness) are measured by θ_{ti}, whereas the item difficulty β_{tj} measures how likely it is to agree with an item. Moreover, individual differences in response styles—which have consistently been found in the literature (e.g., Johnson & Bolt, 2010)—are measured by θ_{mi} and θ_{ei}. Items may also vary in their response-style-related difficulty (e.g., De Jong, Steenkamp, Fox, & Baumgartner, 2008). For example, some items may elicit few extreme responses (i.e., high β_{ej}) or many midpoint responses (i.e., low β_{mj}).

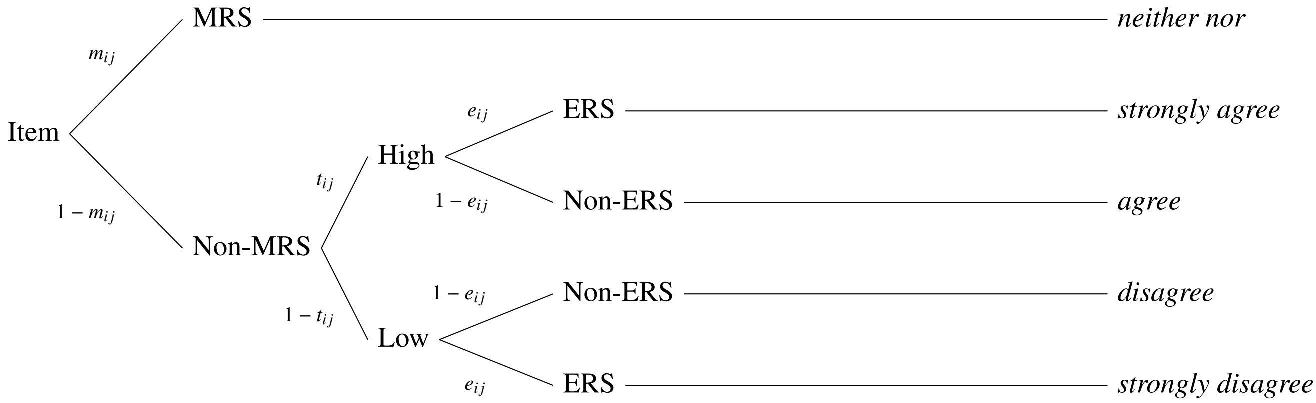

Figure 1. Böckenholt's (2012a) IR-tree model for a 5-point item accounts for midpoint response style (MRS), extreme response style (ERS), and the target trait (High/Low).

According to the tree in Figure 1, the probability of a response is given by multiplying all the probabilities along the corresponding branches (Böckenholt, 2012a). For instance, the probability to strongly agree with an item is given by

$$Pr(x_{ij} = \textit{strongly agree} \mid \boldsymbol{\theta}_i, \boldsymbol{\beta}_j) = (1 - m_{ij}) t_{ij} e_{ij}. \tag{4}$$

It is important to note that neither the Böckenholt Model nor the other tree models discussed in the following make strong assumption about the sequential order of the cognitive processes (e.g., Batchelder & Riefer, 1999). For example, the model equations of the Böckenholt Model can be represented with a tree diagram different from that depicted in Figure 1, for example, one with a "sequence" of t_{ij}, m_{ij}, and e_{ij} (instead of m_{ij}, t_{ij}, and e_{ij}). Hence, tree models make assumptions about the psychological processes but not necessarily about their relative order or temporal sequence.

Böckenholt (2012a), as well as De Boeck and Partchev (2012), proposed to estimate the model using existing maximum likelihood software for IRT models. For this purpose, an observed response is recoded into three binary pseudoitems that correspond to the outcomes of the three latent stages: The first pseudoitem encodes whether the middle category was chosen or not, the second whether the respondent agreed or disagreed, and the third whether an extreme or a moderate response was given (midpoint responses are coded as missing by design on the last two pseudoitems). Based on this recoding, the Böckenholt Model can be fit using software for standard, three-dimensional, binary IRT models. It is important to note that this method is limited to IR-tree models in which each response category is reached by a single processing path (Böckenholt, 2012a; Jeon & De Boeck, 2016).

In summary, the Böckenholt Model assumes three qualitatively distinct processes to account for MRS, ERS, and the target trait. Evidence for the validity of this approach comes from the theoretical derivation of the model based on underlying cognitive processes (Böckenholt, 2012a) as well as from empirical data: For example, the construct validity of the three processes was demonstrated by Hansjörg Plieninger and Meiser (2014) in a study using extraneous style- and content-related criteria. Khorramdel and von Davier (2014) extended the model to questionnaires with multiple domains (Big Five) and were able to show that MRS and ERS are stable across different scales.

Multinomial Processing Tree Models

Tree models are not only used in psychometrics, but also in other fields such as cognitive or social psychology. More specifically, MPT models represent a class of models that allow to disentangle a finite number of qualitatively distinct processes that are assumed to result in identical responses (Erdfelder et al., 2009; Hütter & Klauer, 2016; Riefer & Batchelder, 1988). Recently, MPT models have been generalized to account for hierarchical data structures (Klauer, 2010). Even though MPT and IR-tree models have been developed mostly in isolation, we show in the following that IR-tree models are a special case of hierarchical MPT models. This relationship will be important to guide the development of the proposed ARS model.

MPT models can be illustrated using tree diagrams such as the one shown in Figure 1. In an MPT model, the expected probability for each branch b is obtained by multiplying all parameters ξ_p ($p = 1, \ldots, P$) along the branch b, similar as in an IR-tree model (see Equation 4):

$$Pr(b \mid \boldsymbol{\xi}) = c_b \prod_{p=1}^{P} \xi_p^{v_{bp}} (1 - \xi_p)^{w_{bp}}. \tag{5}$$

Here, v_{bp} and w_{bp} count how often the parameters ξ_p and $(1 - \xi_p)$ occur in branch b, respectively, and c_b represents possible constants due to parameter constraints (Hu & Batchelder, 1994). However, in contrast to IR-tree models or formal trees (undirected acyclic graphs) as defined in mathematical graph theory, MPT models often assume that a category can be reached by more than one branch, because different cognitive processes are assumed to result in the same observable response. For such models, the predicted probability for category x_k ($k = 1, \ldots, K$) is obtained by adding the probabilities of the corresponding branches $b = 1, \ldots, B_k$:

$$Pr(x_k \mid \boldsymbol{\xi}) = \sum_{b=1}^{B_k} P(b \mid \boldsymbol{\xi}). \tag{6}$$

Given these expected category probabilities, the observed response frequencies are assumed to follow a multinomial distribution. To estimate the model parameters ξ_p, they need to be identifiable, which means that identical expected category probabilities $Pr(\boldsymbol{\xi}) = Pr(\boldsymbol{\xi}')$ must imply identical parameter values $\boldsymbol{\xi} = \boldsymbol{\xi}'$ (Batchelder & Riefer, 1999; Schmittmann, Dolan, Raijmakers, & Batchelder, 2010). A necessary (but not sufficient) condition for the identifiability of multinomial models is that the number of parameters does not exceed the number of free categories.

MPT models have often been used under the assumption that observations are independent and identically distributed, or equivalently, that parameters are invariant across persons and items. However, this restrictive assumption of parameter homogeneity has been questioned in recent years, and this was accompanied by the call for models that take heterogeneity of persons and/or items into account (e.g., Rouder & Lu, 2005). Recently, Klauer (2010) and Matzke, Dolan, Batchelder, and Wagenmakers (2015) have developed hierarchical MPT models that include person- and/or item-specific effects, thereby overcoming the need to aggregate the data. Similar to IRT models, hierarchical MPT models assume that the parameters ξ_{pij} are allowed to vary over both persons i and items j. For each person-item combination, the MPT parameters $(\xi_{1ij}, \ldots, \xi_{Pij})$ determine the expected category frequencies as in Equation 6. In addition, the MPT parameters ξ_{pij} are reparameterized using an IRT-like structure with additive person and item effects similar as in IR-tree models (Equations 1-3). More specifically, the probability parameters ξ_{pij} on the interval $[0, 1]$ are first mapped to the real line using, for instance, the probit-link function $\Phi^{-1}(\xi_{pij})$. Then, on the probit scale, the person ability parameter θ_{pi} and the item difficulty parameter β_{pj} are assumed to combine additively,

$$\xi_{pij} = \Phi\left(\theta_{pi} - \beta_{pj}\right). \tag{7}$$

Put differently, each MPT parameter ξ_{pij} is first modeled as the dependent variable of a binary IRT model (i.e., a probit-link IRT or Rasch model). Then, the MPT parameters in Equation 7 are plugged into the MPT model in Equations 5 and 6 separately for each person-item combination. For identification of the θ_{pi} and β_{pj} parameters, a constraint is needed similarly as in standard IRT models (see, e.g., Fox, 2010), for example, person parameters centered at zero. Then, given an identifiable MPT model, the corresponding hierarchical version is also identifiable (Matzke et al., 2015).

This illustrates that the Böckenholt Model can be interpreted as a special case of a hierarchical MPT model, since the model equations of the category probabilities are obtained by multiplying all parameters ξ_p along branch b (Equation 4 and 5). However, estimating IR-tree models based on pseudoitems in general involves “the restriction that each observed response category has a unique path to one of the latent response processes” (Böckenholt, 2012a, p. 667). Thus, models are excluded in which two branches lead to the same category (Jeon & De Boeck, 2016).[2] However, this restriction does not apply to hierarchical MPT models,

[2] Böckenholt (2012b, 2014), and Thissen-Roe and Thissen (2013) presented specific models that do not entail this restriction, and the general framework of hierarchical MPT models subsumes these models. Nevertheless, as noted by Jeon and De Boeck (2016), such models cannot be estimated with the pseudoitem approach, the strategy mainly adopted

where multiple processes may lead to the same outcome with a probability given as the sum of the respective branch probabilities.

Overall, the combination of psychometric measurement models with cognitive process models provides a powerful framework that has received considerable attention in cognitive psychology but not yet in psychometrics. In the present work, we build on the similarity of MPT and IR-tree models to develop a novel, cognitively-inspired model of acquiescence. Thereby, we also want to raise awareness for this modeling approach of *cognitive psychometrics* in general (Riefer, Knapp, Batchelder, Bamber, & Manifold, 2002).

A Hierarchical MPT Model of Acquiescence

The Böckenholt Model is limited to two response styles, namely, ERS and MRS. We propose an extension that also takes ARS into account and that can be implemented as a hierarchical MPT model. The proposed *Acquiescence Model* builds on the basis of the Böckenholt Model and adds an additional processing stage to it. As shown in Figure 2, respondents are presented with an item and may enter a "non-acquiescent stage" with probability $1 - a_{ij}$, which leads to the original predictions of the Böckenholt Model (see Figure 1). However, with probability a_{ij}, respondents enter an "acquiescent stage" that always results in affirmative responses irrespective of the coding direction of the item and irrespective of the lower part of the tree. In other words, the Acquiescence Model assumes two distinct processes that lead to agreement with the items—respondents either agree because of the item's content (target trait) or merely due to a general tendency to provide affirmative responses (ARS). Analogously to the other MPT parameters, the ARS parameter is decomposed as follows:

$$a_{ij} = \Phi(\theta_{ai} - \beta_{aj}). \tag{8}$$

Respondents may differ in their ARS-level, which is captured by θ_{ai}, and items may elicit ARS responses to different degrees, which is captured by β_{aj}.

Five-point items have two affirmative categories, namely, a moderate (i.e., *agree*) and an extreme one (i.e., *strongly agree*). Therefore, an additional MPT parameter e^*_{ij} is necessary to model the probability of extreme responses conditional on acquiescence. The most flexible model entails a reparameterization of this parameter as above, namely, $e^*_{ij} = \Phi(\theta_{e^*i} - \beta_{e^*j})$. However, we assume that respondents have a general tendency towards extreme (or moderate) responses, which does not depend on acquiescence. Thus, we set the person parameters equal across branches, namely, $\theta_{e^*i} = \theta_{ei}$, which implies that respondents high on ERS do not only prefer extreme categories when giving a content-related response, but do so similarly in case of ARS-responding. Furthermore, we constrain all respective item parameters to be equal, namely, $\beta_{e^*j} = \beta_{e^*}$, and the implications of relaxing this constraint will be discussed in the empirical example on page 17. Taken together, the equation for the last MPT parameter is then $e^*_{ij} = \Phi(\theta_{ei} - \beta_{e^*})$. Note that these two constraints are not necessary for identification (see below) but render the model more parsimonious, which in turn facilitates parameter estimation.

The complete model for regular items (reg.) is then defined by the following set of equations:

$$\begin{aligned} &Pr(x_{ij} = \textit{strongly agree} \mid \text{reg.}) \\ &\quad = a_{ij}e^*_{ij} + (1 - a_{ij})(1 - m_{ij})t_{ij}e_{ij} \end{aligned} \tag{9}$$

$$\begin{aligned} &Pr(x_{ij} = \textit{agree} \mid \text{reg.}) \\ &\quad = a_{ij}(1 - e^*_{ij}) + (1 - a_{ij})(1 - m_{ij})t_{ij}(1 - e_{ij}) \end{aligned} \tag{10}$$

$$\begin{aligned} &Pr(x_{ij} = \textit{neither nor} \mid \text{reg.}) \\ &\quad = (1 - a_{ij})m_{ij} \end{aligned} \tag{11}$$

$$\begin{aligned} &Pr(x_{ij} = \textit{disagree} \mid \text{reg.}) \\ &\quad = (1 - a_{ij})(1 - m_{ij})(1 - t_{ij})(1 - e_{ij}) \end{aligned} \tag{12}$$

$$\begin{aligned} &Pr(x_{ij} = \textit{strongly disagree} \mid \text{reg.}) \\ &\quad = (1 - a_{ij})(1 - m_{ij})(1 - t_{ij})e_{ij}. \end{aligned} \tag{13}$$

Importantly, any ARS model requires both regular and reversed items in order to disentangle ARS and the target trait. The same condition also applies to the Acquiescence Model, which therefore comprises two distinct processing trees: The first tree concerns regular items and is shown in Figure 2, whereas the second tree concerns reversed items and is not shown due to space considerations. The reversed tree is identical to that in Figure 2 with a single exception: For reversed

in the IR-tree literature (Böckenholt, 2012a; De Boeck & Partchev, 2012).

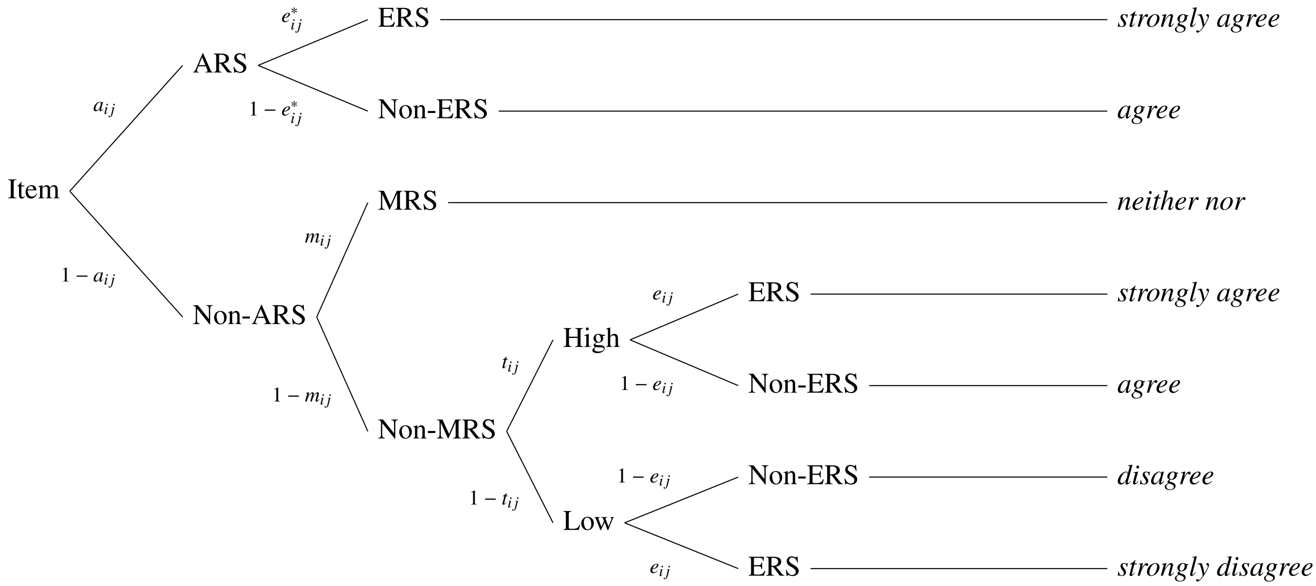

Figure 2. The Acquiescence Model for a regular 5-point item accounts for midpoint response style (MRS), acquiescence response style (ARS), and extreme response style (ERS) besides the target trait (High/Low). Note that multiple branches lead to agreement thereby indicating a mixture of the target trait and the ARS distribution.

items, the high target-trait stage eventually leads to disagreement, and the low target-trait stage eventually leads to agreement.[3]

Concerning the identifiability of the proposed Acquiescence Model, the necessary condition for multinomial models discussed above is satisfied, since the model is comprised of only five free parameters (m, e, a, e^*, t) to model 8 non-redundant category probabilities ($5 - 1$ for regular as well as reversed items). To check a sufficient condition of identifiability, we used the computer algebra system described by Schmittmann et al. (2010): This showed that identical expected category probabilities $Pr(\boldsymbol{\xi}) = Pr(\boldsymbol{\xi}')$ indeed imply identical parameter values $\boldsymbol{\xi} = \boldsymbol{\xi}'$, which means that the model parameters $\boldsymbol{\xi} = (m_{ij}, e_{ij}, a_{ij}, e^*_{ij}, t_{ij})$ are identifiable for specific person-item combinations. Furthermore, the IRT parameters θ_{pi} and β_{pj} that are used to reparameterize the MPT parameters are rendered identifiable by centering the hyperpriors for the person parameters $\boldsymbol{\theta}_i$ at $\mathbf{0}$ (see below; Fox, 2010, p. 86). Overall, this shows that the Acquiescence Model in Figure 2 is identifiable even without the constraints introduced above. However, it should be noted that, in a given data set, specific parameters may be poorly identified empirically leading to high uncertainty. For example, estimating parameter e^*_{ij} will be more difficult the lower the probability a_{ij}, because the former is defined conditionally on the latter (see Figure 2). Likewise, disentangling the parameters a_{ij} and t_{ij} will be more difficult the fewer reversed items are used.

The presented model has some notable special cases as well as straightforward extensions. First, the Acquiescence Model reduces to the Böckenholt Model if $a = 0$ (i.e., if $(\theta_a - \beta_a) \rightarrow -\infty$). Substantively, this is the case, for instance, if respondents are very low on acquiescence. Furthermore, the model reduces to a Rasch model with a probit link if the number of categories is two and if the number of parameters $P = 1$.

Second, the model can be extended to more than one content domain requiring a model with multiple target traits t_d ($t_d = t_1, \ldots, t_D$). In this case, each domain is modeled using separate trees, equations, and θ_{t_d}-parameters, which all increase in number by the factor D. The response style parameters should be set equal across domains mirroring the assumption that response styles are stable across content domains (Danner et al., 2015; Khorramdel & von Davier, 2014).

[3]Likewise, the complete model is expressed by two sets of equations. Equations 9 to 13 hold for regular items, and five additional equations are needed for reversed items. These equations mirror Equations 9 to 13 with the exceptions that $(1 - t_{ij})$ is replaced by t_{ij} and t_{ij} is replaced by $(1 - t_{ij})$.

Apart from that, the IRT part in Equation 7 may take on more complex forms, for example, by including an item-discrimination parameter (e.g., Jeon & De Boeck, 2016; Khorramdel & von Davier, 2014).

Mixture Versus Shift Models for Acquiescence

The most prominent alternative ARS model was developed within the framework of confirmatory factor analysis. A two-factor model (random-intercept model) is specified with (a) a target-trait factor θ^*_{ti} with loadings λ_j that are positive for regular items and negative for reversed items and (b) an ARS factor θ^*_{ai} with loadings fixed to 1 (e.g., Billiet & McClendon, 2000; Maydeu-Olivares & Coffman, 2006). The model is described by the following generic equation[4], adapted to our notation:

$$f(x_{ij}) = \lambda_j \theta^*_{ti} + \theta^*_{ai} - \beta_j. \tag{14}$$

Note that starred versions of θ are used to distinguish the person parameters from those used in the Acquiescence Model above. The two factors in Equation 14 operate additively on the latent scale. Hence, this makes the model a *shift model* that assumes that the tendency to agree or disagree with an item follows an up- or downwards shift determined by the ARS factor (see Figure 3 for an example). Note further that different variants of this model exist (e.g., Billiet & McClendon, 2000; Falk & Cai, 2016; Ferrando, Morales-Vives, & Lorenzo-Seva, 2016; Kam & Zhou, 2015; Maydeu-Olivares & Coffman, 2006). However, the basic idea of two additive, compensatory factors is common to all these approaches.

It is worth noting the following theoretical implications of this shift model. First, the model implies that acquiescence and disacquiescence are opposite poles of a single dimension. Hence, high ARS-values θ^*_{ai} predict a shift towards agreement (see Figure 3F), low values predict a shift towards disagreement (Figure 3D), and intermediate values imply the absence of an ARS effect (Figure 3E). Second, the model is a *compensatory model* (e.g., Babcock, 2011; Bolt & Lall, 2003), because, for instance, high ARS-levels and low target-trait levels outweigh each other. Third, due to this compensatory nature of the shift model, an ARS effect may result in a shift from a *strongly disagree* response to a *disagree* response. In such a case, however, ARS actually predicts disagreement, despite the theoretical definition of acquiescence as the tendency to prefer agree responses.

Contrary to this shift model, the novel Acquiescence Model is a *mixture model* of acquiescence. Agreement with an item may emerge from two distinct processes, namely, either from the target trait or from ARS (see Figure 2). Hence, ARS does not affect the observable distribution of response frequencies by additive shifts. Instead, the model predicts a mixture of two underlying distributions of ARS and content-related responding with mixing probabilities a_{ij} and $(1-a_{ij})$, respectively (see Equations 9 to 13). Figure 3 illustrates three implications, in which this mixture model qualitatively differs from the shift model: First, the opposite "pole" of ARS is the absence of ARS and not disacquiescence as in the shift model. That is, probabilities a_{ij} close to 0 imply the absence of an ARS effect (Figure 3A). In such a case, the Acquiescence Model (Figure 2) reduces to the Böckenholt Model (Figure 1) because the ARS-branch is never reached. Second, the Acquiescence Model is *non-compensatory* in nature. Substantively, high levels of acquiescence result in entering the ARS branch of the processing tree in which the target-trait parameters θ_{ti} and β_{tj} play no role and cannot compensate for high ARS levels. Third, an increase in a_{ij} increases the probabilities for the two *agree*-categories and decreases the probabilities for the three non-agree categories. This means that an ARS effect may shift disagreement (predicted by t_{ij}) to agreement, but—in contrast to the shift model—a shift from a *strongly disagree* response to a *disagree* response is impossible.

Besides these qualitative differences, the shift and the mixture approach also share important properties. In both models, higher ARS-levels increase the probability of affirmative responses. Furthermore, high ARS may lead to agreement with both regular and reversed items in both models. Moreover, it is possible to formulate a shift model within the IR-tree framework, and this is described in Appendix A.

A direct comparison of the shift and the mixture account is difficult to make on theoretical grounds, because definitions of ARS are not very precise and can be interpreted in either way. Some measures of ARS

[4]Please refer to the cited references for details. Even though irrelevant to the discussion herein, note that often a linear model without a link function and without category-specific item parameters is used.

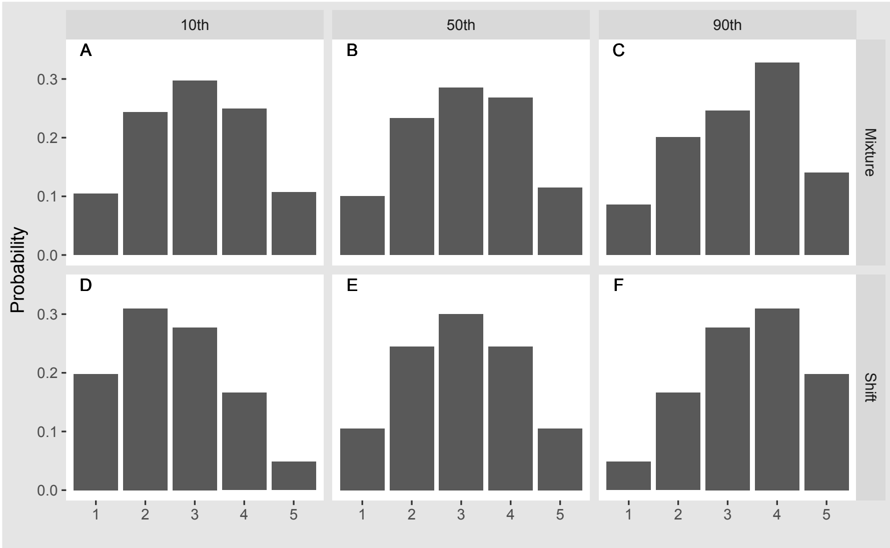

Figure 3. The effect of a mixture versus a shift model on the predicted response distributions of a hypothetical 5-point item. Displayed are the probabilities for respondents at the 10th, 50th, and 90th percentile of the ARS distribution.

are more in line with a shift model, for example, when the mean across both regular and reversed items is used (e.g., Couch & Keniston, 1960); and some measures of ARS are more in line with a mixture model, for example, when the number of affirmative responses is counted (e.g., Billiet & McClendon, 2000). Furthermore, from the perspective of a shift model, measures of acquiescence and disacquiescence should show strong, negative correlations. However, values of .28 (Weijters et al., 2010b) or −.16 (Baumgartner & Steenkamp, 2001) are not consistent with this prediction. In summary, the mixture approach allows to shed new light on the well-known phenomenon of acquiescence. However, the aim of the present work is not to rule out the shift model, but rather to point out qualitative differences in conceptions and models of ARS, which may in part be responsible for mixed empirical findings as discussed above.

Bayesian IRT: Priors and Estimation

We adapted the hierarchal priors and weakly informative hyperpriors proposed by Matzke et al. (2015), which are similar to those in standard Bayesian IRT models with random person and random item effects (see also Fox, 2010). For the Böckenholt Model, we checked that these priors resulted in parameter estimates closely resembling those based on the maximum-likelihood analysis proposed by Böckenholt (2012a). For the person parameters, we assume a centered, multivariate normal distribution,

$$\boldsymbol{\theta}_i \sim \text{Multivariate-Normal}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{15}$$

with a covariance matrix $\boldsymbol{\Sigma}$ estimated from the data. In contrast, the item parameters have independent, univariate normal priors. To allow for the possibility to define different hyperpriors for response styles and the target trait(s) (indexed by $p = (e, m, a, t_1, \ldots, t_D)$), the item parameters β are partitioned into an additive combination of mean μ and centered differences δ:

$$\beta_{pj} = \mu_p + \delta_{pj}. \tag{16}$$

Note that such a decomposition also improves convergence when estimating the parameters (Matzke et al., 2015). Based on this parameterization, the following prior and hyperprior distributions were used:

$$\mu_p \sim \text{Truncated Normal}(0, 1, -5, 5) \tag{17}$$

$$\beta_{e^*} \sim \text{Truncated Normal}(0, 1, -5, 5) \tag{18}$$

$$\delta_{pj} \sim \text{Truncated Normal}(0, \sigma^2_{\beta_p}, -5, 5) \tag{19}$$

$$\sigma^2_{\beta_p} \sim \text{Inverse-Gamma}(1, 1) \tag{20}$$

$$\boldsymbol{\Sigma} \sim \text{Scaled Inverse-Wishart}(\boldsymbol{I}_P, df = P + 1, \tau_p) \tag{21}$$

$$\tau_p \sim \text{Uniform}(0, 100). \tag{22}$$

The priors for the item parameters μ_p, β_{e^*}, and δ_{pj} were truncated to aid faster convergence. Given that latent probit values larger than 5 result in negligible response probabilities smaller than $3 \cdot 10^{-7}$, this does not constrain or inform parameter estimation substantially (but this restriction can also be dropped). The inverse-Wishart prior for the covariance matrix $\mathbf{\Sigma}$ in Equation 21 (parameterized by $P+1$ degrees of freedom and the P-dimensional identity matrix $\boldsymbol{I}_P$) implies marginal uniform priors on the correlations. Moreover, we used a scaled version of the inverse-Wishart prior with the scale parameters τ_p (see Equation 22) that maintains this property but is less restrictive with respect to the variances in $\mathbf{\Sigma}$ (Gelman et al., 2014, p. 74).

Because an analytical solution for the full posterior is not available, the model is estimated by approximating the posterior distribution by Markov chain Monte Carlo (MCMC) sampling using JAGS (Denwood, 2016; Plummer, 2003), a popular software for Gibbs sampling. To cross-check our results, we also implemented the model in Stan (Carpenter et al., 2017), a more recent software package that draws posterior samples based on adaptive Hamiltonian Monte Carlo (Hoffman & Gelman, 2014), a more efficient sampling scheme that often reduces auto-correlation. R (R Core Team, 2016) was used as a front-end in both cases, and our R package mpt2irt for parameter estimation is available from https://github.com/hplieninger/mpt2irt/.

Simulation Studies

We performed two simulation studies (a) to investigate the recovery of core parameters of the Acquiescence Model and (b) to compare the Böckenholt Model and the Acquiescence Model when fit to data generated from each of both.

The simulations were summarized using the posterior medians $\hat{\pi}_p$ as estimates for the true parameters π_p (where π_p stands for the person and item IRT parameters θ_{pi} or β_{pj}). In each replication and for each parameter π_p, three measures were calculated across persons or items, namely the correlation $r_{\hat{\pi}_p,\pi_p}$, the mean bias (i.e., $\text{Mean}(\hat{\pi}_p - \pi_p)$), and the RMSE (i.e., $\sqrt{\text{Mean}(\hat{\pi}_p - \pi_p)^2}$). Below, we report summaries of these three measures across replications.

Study 1: Parameter Recovery

Method. Data were generated from the Acquiescence Model for 1,000 persons. A condition with 20 and a condition with 40 items was realized with half of the items being reversed. In each replication, the person parameters were drawn from a centered multivariate normal distribution, $\mathbf{\Theta} \sim \text{MVN}(\mathbf{0}, \mathbf{\Sigma})$. The variances in $\mathbf{\Sigma}$ were set to $\sigma^2_{\theta_m} = \sigma^2_{\theta_e} = \sigma^2_{\theta_a} = 0.33$ and $\sigma^2_{\theta_t} = 1.00$ mirroring the fact that content-related variance is usually larger than response-style-related variance in empirical data (e.g., Billiet & McClendon, 2000). The covariances in $\mathbf{\Sigma}$ were drawn from a Wishart distribution with $df = 50$ and the scale matrix

$$\mathbf{\Sigma}^* = \begin{bmatrix} 1 & -.2 & 0 & 0 \\ -.2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{23}$$

which mirrors the empirical finding of a negative correlation between MRS and ERS (with 90 % of the simulated correlations in the range $[-.41, 0.04]$), but small correlations otherwise (90 % in the interval $[-.23, .23]$). Furthermore, the item parameters in each replication were drawn from independent truncated normal distributions with $\mu_{\beta_m} = \mu_{\beta_e} = \mu_{\beta_{e^*}} = \Phi^{-1}(.70)$, $\mu_{\beta_a} = \Phi^{-1}(.95)$, and $\mu_{\beta_t} = \Phi^{-1}(.50)$, and with $\sigma^2_{\beta_m} = \sigma^2_{\beta_e} = \sigma^2_{\beta_a} = \sigma^2_{\beta_{e^*}} = 0.10$ and $\sigma^2_{\beta_t} = 0.50$.[5] This implies, for example, that the expected probabilities for an average person with $\boldsymbol{\theta}_i = \mathbf{0}$ were .05 versus .95 for an ARS versus non-ARS response, respectively, and .30 versus .70 for an MRS versus non-MRS response, respectively (conditional on a non-ARS response). Furthermore, the item parameters β_{tj} of the target trait were simulated with a larger variance than those for the three response style processes.

We generated 200 data sets for each of the two conditions with 20 and 40 items, respectively, and fit the Acquiescence Model using Stan. Posterior samples were obtained from three independent chains with 1,000 iterations each, of which the first 500 were discarded. In preliminary analyses, these values were fine-tuned to balance computation time and precision.

[5]These values are partly based on theoretical considerations (e.g., the ARS prevalence should be very low) and partly based on preliminary analyses of empirical data sets.

Results. Concerning the correlations of data-generating and fitted parameters, recovery was better for the item than for the person parameters, for obvious reasons: The former were informed by 1,000 persons whereas the latter were informed by only 20 and 40 items, respectively (see Figure 4A). Recovery generally improved when using more items, which is an indication of the consistency of the estimation procedure. With respect to the β-parameters, recovery was best for the β-parameters for MRS and ERS, because these two processes can directly be inferred from observed responses. The tree in Figure 2 illustrates this property of the Acquiescence Model, that midpoint and extreme responses uniquely emerge from MRS and ERS, respectively. In contrast, affirmative responses are either due to the target trait t_{ij} or due to acquiescence a_{ij}, which diminishes the precision of the estimates for β_t and β_a. Moreover, the generated ARS item parameters β_a were much larger than the other parameters, mirroring a low prevalence of acquiescent behavior, and this additionally reduces the precision of parameter estimates, which is further illustrated in Appendix C. With respect to the θ-parameters, a similar pattern as for the β-parameters was observed, with the exception that recovery of θ_t was best due to the fact that the variance of this dimension was larger than that of the three other dimensions.

Concerning the mean bias for MRS, ERS, and the target trait, the estimates for both the item parameters β_{pj} and the person parameters θ_{pi} were unbiased (see Figure 4B). The ARS parameters, however, were overestimated due to the low prevalence of ARS (see also Appendix C). For the β_a-parameters, this guards against a type I error of incorrectly classifying an item as suspicious (i.e., being susceptible to ARS) at the cost of statistical power. Bias was less severe for the θ_a-parameters, and this bias was in particular caused by upwards-shrinkage towards zero for persons low on ARS (because it is hard to tell from only a few items whether such a person has a θ_a-parameter of, say, -0.7 or -0.5). The results for RMSE mirrored those of the correlations reported above. RMSE was generally smaller for the item parameters β than for the person parameters θ, and generally smaller with 40 compared to 20 items (see Figure 4C). Aside from the core parameters for MRS, ERS, ARS, and the target trait, recovery of the single parameter β_{e^*} that measures extreme responding conditional on ARS was comparable to that of the β_e-parameters with respect to correlation, bias, and RMSE.

Study 2: Model Recovery

Method. In the second study, 200 data sets were generated from the Böckenholt Model and 200 from the Acquiescence Model. The Böckenholt as well as the Acquiescence Model were fit to each data set. Each data set included 250 persons and 20 items (half of which were reversed). The data-generation procedure was identical to that of Study 1 with the obvious exception that the a_{ij}-parameter is absent in the Böckenholt Model (and the covariance matrix $\mathbf{\Sigma}$ reduces to three dimensions).

The models were fit using JAGS which facilitates the computation of the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002), a model-selection criterion that trades-off goodness of fit (i.e., minus the expected deviance) against model complexity (i.e., the effective number of free parameters). To select a model, DIC is computed for each of the competing models and the one with the smallest DIC value is selected. Again, we sampled three chains each with 500 retained iterations, but the computations in JAGS required a longer burn-in phase of 1,000 iterations and a thinning factor of 10 (i.e., the number of iterations was actually 10 times larger, but only every 10th iteration was retained).

Results. At first, we consider the condition in which data were generated from the Acquiescence Model. As expected, DIC was smaller for the Acquiescence compared to the Böckenholt Model in 82 % of the cases, indicating that the two models can, with a realistic set-up of 250 persons and 20 item, in principle be discriminated using this model-selection criterion. The preference of DIC towards the Acquiescence Model increased in data sets in which acquiescence was more prevalent: ΔDIC was lower (i.e., more favorable of the Acquiescence Model) the larger the ARS variance $\sigma^2_{\theta_a}$ was ($r = -.21$), and ΔDIC was lower the lower the mean of the β_a-parameters was (thus eliciting ARS responses more easily; $r = .31$).

A closer look at the estimated parameters shows that the item parameters, compared to the person parameters, were more affected when fitting the Böckenholt Model instead of the data-generating Acquiescence

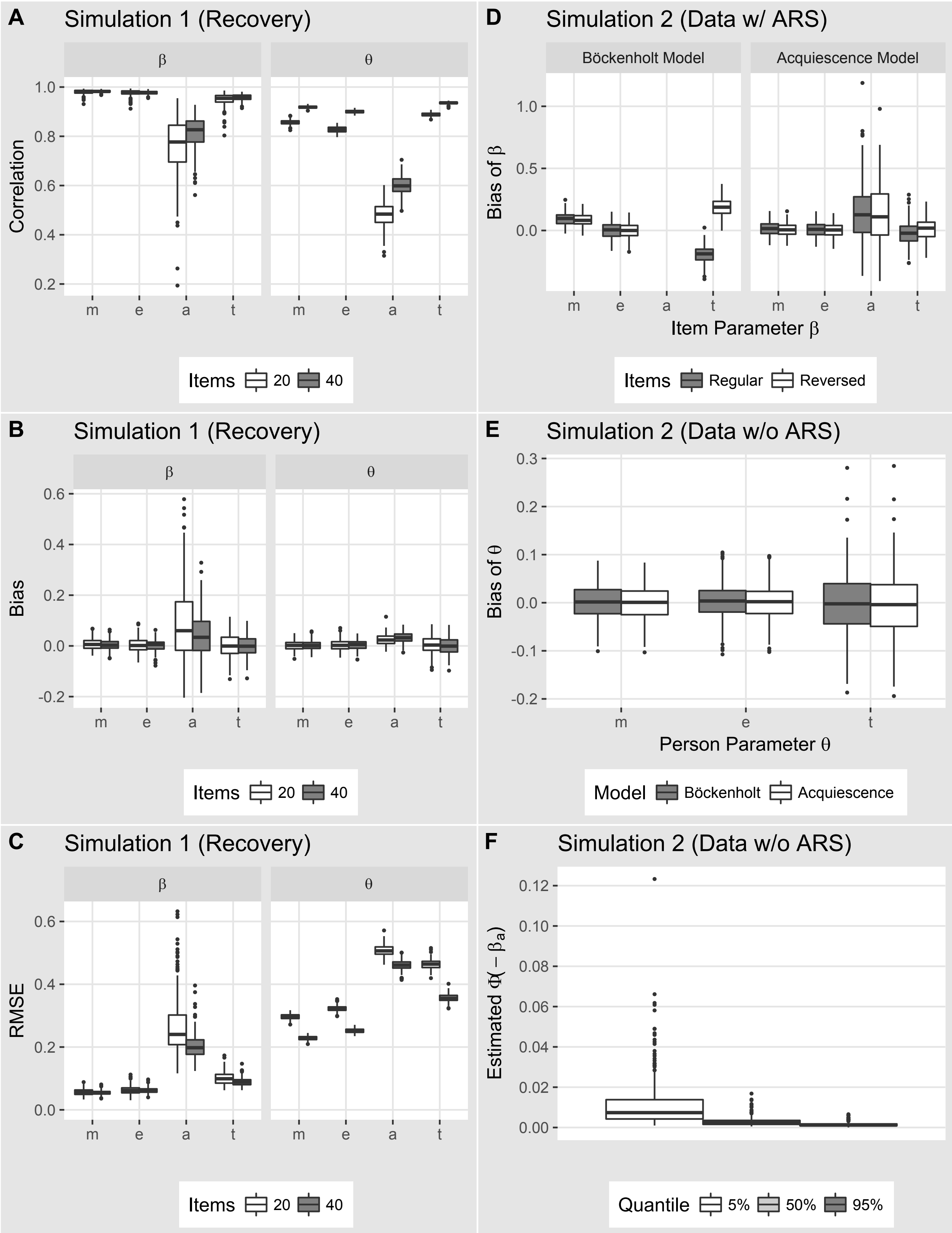

Figure 4. Boxplots in panels A to C display the results of Study 1, where the Acquiescence Model was the data-generating and the fitted model. Boxplots in panels D to F display the results of Study 2, where the data-generating model was either the Acquiescence Model (Panel D) or the Böckenholt Model (Panels E and F).

Model, an effect that was largest for the target-trait difficulties β_t. Average recovery of these parameters was considerably better when fitting the correctly specified Acquiescence Model ($r_{\hat{\beta}_t,\beta_t} = .86$, $RMSE = 0.16$) instead of the misspecified Böckenholt Model ($r_{\hat{\beta}_t,\beta_t} = .69$, $RMSE = 0.25$). This was due to the fact that the Böckenholt Model does not take into account the confounding of the target trait t_{ij} and ARS a_{ij}; thus, the β_t-parameters were biased, namely, underestimated for regular items and overestimated for reversed items (see Figure 4D). Furthermore, the β_m-parameters were biased if ARS was not taken into account. The person parameters estimated from the Böckenholt Model were—even though inferior to those from the Acquiescence Model—less affected than the item parameters, and therefore do not receive further discussion herein.

Next, we consider the condition in which the data were generated using the Böckenholt Model, that is, without acquiescence. Since this model is nested within the Acquiescence Model, the simulated data are still compatible with the latter model. However, model selection should prefer the Böckenholt Model because of its smaller complexity due to the absence of any (here superfluous) ARS parameters. This was indeed the case, DIC was smaller for the Böckenholt compared to the Acquiescence Model in 90 % of the cases indicating that this criterion again allowed to discriminate between the two models. With respect to recovery of the person parameters, the two models performed equally well, and the Böckenholt Model as well as the overly complex, misspecified Acquiescence Model resulted in unbiased estimates (see Figure 4E). However, the Böckenholt Model was slightly more accurate in estimating the item parameters, especially the target-trait difficulties β_t.

Of special interest were the estimates for the ARS-parameters when fitting the unnecessarily complex Acquiescence Model to data generated without acquiescence. Essentially, the simulation indicated that the Acquiescence Model empirically reduced to the Böckenholt Model in many respects. First, the estimates for the item parameters β_a were very large mirroring a very low prevalence of ARS. Figure 4F illustrates that, for an average person, the probability to give an ARS response to the item at the median of the 20 items was below 1 % in almost all replications. Second, the variance $\sigma^2_{\theta_a}$ was estimated to be .14 on average, whereas the data-generating variances of MRS and ERS ($\sigma^2_{\theta_m} = \sigma^2_{\theta_e} = .33$) were properly recovered in both models. Taken together, if the more complex Acquiescence Model is fit to a data set with absolutely no acquiescence, DIC is likely to indicate a preference for the more parsimonious Böckenholt model and estimates for the ARS parameters β_a and $\sigma^2_{\theta_a}$ will lead to the correct conclusion that responses were not affected by ARS. In such a case, the estimates for the remaining parameters (especially the item parameters) from the Böckenholt Model are expected to be slightly more precise and are thus preferred when drawing substantive conclusions.

Empirical Study

In the following example, we demonstrate the application of the proposed model with empirical data. First, we fit models for a single content domain followed by models for six personality factors.

Method

We re-analyzed data by Moshagen, Hilbig, and Zettler (2014), who investigated the factorial structure and psychometric quality of the German version of the HEXACO personality inventory. The revised HEXACO personality inventory (HEXACO-PI-R; K. Lee & Ashton, 2016) is comprised of 96 five-point items (*strongly disagree* to *strongly agree*) to measure six scales—namely, honesty-humility (H), emotionality (E), extraversion (X), agreeableness (A), conscientiousness (C), and openness to experience (O)—with the corresponding items presented in alternating order. Within each scale, between seven and ten of the 16 items were reversed with 48 reversed items in total. Our reanalysis is based on the second sample of Moshagen et al. (2014) that includes 1,012 university students, of which we drew a random subsample of 500 respondents. The remaining persons were used for cross-validating the estimated item parameters (see Appendix B).

Two response style models, the Böckenholt and the Acquiescence Model, were fit using Stan (Carpenter et al., 2017) by sampling six chains with 1,500 iterations each, of which the first 500 were discarded (i.e., 6,000 retained iterations in total). The number of retained iterations was quadrupled in comparison to the simulation studies, because a good approximation of the posterior was even more important herein. Convergence of the sampling procedure is graphically illustrated in Appendix D. Again, the posterior distributions were summarized using their medians as estimates as well as 95 % posterior intervals reported in brackets. Note that, for ease of interpretation, the raw item parameters β were transformed to probabilities $\Phi(0 - \beta)$, that is, the probability for an average person with $\boldsymbol{\theta}_i = \mathbf{0}$ to pass the item's threshold. For example, an item parameter $\beta_m = 1$ can be expressed as $\Phi(-1) = .16$: Thus, the probability for an MRS response is 16 % for an average person.

Results

Response style models for a single content domain. We first fit a series of models to items from a single content domain, namely, honesty-humility, which refers to individual differences in treating "others fairly even when one could successfully exploit them" (K. Lee & Ashton, 2016, p. 2). The scale is comprised of six regular and 10 reversed items. Initially, we compared the novel Bayesian implementation of the Böckenholt Model to the previously used maximum-likelihood estimation that is based on binary pseudoitems (Böckenholt, 2012a; De Boeck & Partchev, 2012). Both the item and the person parameters were virtually identical with correlations above .99 and a mean bias of almost zero. Then, the Acquiescence Model was fit to the data, which resulted in a substantial improvement with DIC of 20,167 compared to 20,296 for the Böckenholt Model.[6]

With respect to the item parameters, most variability was observed for β_t (target-trait difficulty) even though most items were rather easy (i.e., easily eliciting responses indicative of high honesty-humility) with a mean $\Phi(-\mu_{\beta_t})$ of .87 [.77, .93]. The transformed MRS item parameters $\Phi(-\beta_m)$ ranged from .15 to .34 with a mean of .22 [.16, .29] indicating that (given a non-ARS response) the middle category was on average chosen with a probability of 22 %. The ERS difficulties $\Phi(-\beta_e)$ ranged from .14 to .48 with a mean of .30 [.22, .39] indicating that—when choosing between an extreme and a moderate category—an extreme response was on average chosen with a probability of 30 %. Of most interest, the novel ARS difficulties $\Phi(-\beta_a)$ were estimated to be considerably low ranging from .01 to .16 with a mean of .04 [.02, .08]. Substantially, this implies that an ARS response was on average expected in only 4 % of the cases. This prevalence might seem rather small at first sight, but it was expected given the model definition. Essentially, ARS responses are assumed to be independent of content-related response processes and should therefore occur infrequently when using both dependable samples and solid psychometric questionnaires. However, a few items showed higher levels of acquiescence, an observation that is discussed further below.

Besides the item parameters, the person parameters allow for additional insights. As expected, response style variance was smaller than target-trait variance with estimates of $\sigma^2_{\theta_m} = 0.21$ [0.16, 0.26], $\sigma^2_{\theta_e} = .49$ [0.41, 0.60], $\sigma^2_{\theta_a} = 0.32$ [0.17, 0.51], and $\sigma^2_{\theta_t} = 1.13$ [0.88, 1.46]. Note that the variance of acquiescence, which is ignored in the Böckenholt Model, was estimated to be larger than that of MRS, thereby indicating the importance of ARS. Correlations between different response styles were estimated to be rather small with the exception of MRS and ERS, which correlated negatively at −.53 [−.64, −.41], which is a typical finding.

Response style models for all six HEXACO factors. The target trait(s) and response styles are in general more easily disentangled when using content-heterogeneous items as found in multidimensional questionnaires, thereby facilitating the detection and estimation of response styles (e.g., Khorramdel & von Davier, 2014; Weijters et al., 2010b). The same holds for the Böckenholt and the Acquiescence Model, which we fit to multiple domains simultaneously with the constraint that the response style parameters θ_m, θ_e, and θ_a are identical across all items. Thereby, precision of the corresponding estimates is expected to increase. A shortcoming of standard estimation techniques (e.g., the EM algorithm; Dempster, Laird, & Rubin, 1977) is that such high-dimensional models become computationally intractable when the number of dimensions becomes large, say larger than four (e.g., Fox, 2010). However, this limitation does not apply to Bayesian implementations. Therefore, we were able to estimate a 9-dimensional version of the Acquiescence Model comprised of six target traits and three response styles.

In terms of fit, the Acquiescence Model was superior compared to the Böckenholt Model with DIC values of 122,348 and 123,024, respectively, indicating

[6]The DIC was always estimated separately using JAGS with a thinning factor of 10.

the importance of taking ARS into account. Model fit was further evaluated by means of posterior predictive checks reported in Appendix B. The estimated item parameters of the Acquiescence Model are displayed in Figure 5. Across the six HEXACO scales, the content-related item parameters were most variable (with variances ranging from $\sigma^2_{\beta_\mathrm{H}} = 0.51$ to $\sigma^2_{\beta_\mathrm{A}} = 1.16$) and rather easy (with means ranging from $\Phi(-\mu_{\beta_\mathrm{A}}) = .51$ to $\Phi(-\mu_{\beta_\mathrm{H}}) = .87$). In contrast, the MRS and ERS parameters were much more homogeneous with $\sigma^2_{\beta_m} = 0.07$ $[0.06, 0.10]$ and $\sigma^2_{\beta_e} = 0.13$ $[0.10, 0.18]$. Moreover, the thresholds for these response styles showed low mean probabilities of $\Phi(-\mu_{\beta_m}) = .25$ $[.23, .27]$ and $\Phi(-\mu_{\beta_e}) = .26$ $[.23, .29]$. Substantively, this implies that midpoint and extreme responses were on average given with conditional probabilities of 25 % and 26 %, respectively. The ARS parameters β_a, which were of particular interest here, were rather difficult with $\Phi(-\mu_{\beta_a}) = .03$ $[.02, .04]$ indicating that ARS responses were unlikely for most of the items. However, a few items stood out, for example, $\Phi(-\beta_{a,67}) = .14$ $[.09, .19]$. For this item, a person with an average ARS-level (i.e., $\theta_{ai} = 0$) has a probability of 14 % of agreeing with this item irrespective of his or her standing on the target trait $\theta_{t_d i}$. For a person with an ARS-level θ_{ai} one standard deviation above or below the mean, the probability of an ARS response changes to 25 % and 7 %, respectively. The probability corresponding to parameter β_{e^*} (i.e., the ERS-difficulty conditional on ARS) was rather low with $\Phi(-\beta_{e^*}) = .03$ $[.02, .05]$, thereby indicating that the *agree*-category was preferred over the *strongly agree*-category in the ARS-branch.

Regarding the person parameters, most variability was observed with respect to the target trait (with variances ranging from $\sigma^2_{\theta_\mathrm{O}} = .65$ $[0.53, 0.81]$ to $\sigma^2_{\theta_\mathrm{H}} = 1.20$ $[0.96, 1.52]$), compared to smaller variances for the response-style-related processes (i.e., $\sigma^2_{\theta_m} = 0.06$ $[0.05, 0.08]$, $\sigma^2_{\theta_e} = 0.26$ $[0.23, 0.31]$, and $\sigma^2_{\theta_a} = 0.16$ $[0.11, 0.23]$). Similarly as in the single-domain analysis, MRS was negatively correlated with ERS and with ARS, in contrast to a positive correlation between ERS and ARS (see Table 1). Importantly, the content–style correlations were rather small with a mean absolute correlation of .09 (see Table 1). Larger values were observed for the relationship of ARS with both honesty-humility ($r = -.36$ $[-.51, -.20]$) and conscientiousness ($r = -.24$ $[-.40, -.08]$). Even though research on the

Table 1
Estimated Latent Correlations and Variances of the Acquiescence Model for all 96 HEXACO Items

	Response styles			Personality traits					
	MRS	ERS	ARS	H	E	X	A	C	O
MRS	0.06	[−.49, −.30]	[−.45, −.14]	[−.20, .04]	[−.08, .15]	[−.28, −.05]	[−.09, .14]	[−.04, .20]	[−.26, −.03]
ERS	−.40	0.26	[.27, .53]	[−.14, .09]	[−.06, .15]	[−.12, .09]	[−.12, .09]	[−.14, .08]	[−.08, .14]
ARS	−.30	.41	0.16	[−.51, −.20]	[−.23, .08]	[−.29, .04]	[−.27, .04]	[−.40, −.08]	[−.21, .12]
H	−.08	−.02	−.36	1.20	[−.05, .18]	[−.15, .09]	[.23, .45]	[−.02, .23]	[.04, .28]
E	.03	.05	−.07	.07	0.77	[−.14, .10]	[−.34, −.12]	[−.07, .17]	[−.08, .16]
X	−.17	−.01	−.13	−.03	−.02	1.10	[−.11, .12]	[.07, .31]	[.03, .28]
A	.03	−.02	−.12	.35	−.23	.01	0.96	[−.15, .10]	[−.07, .17]
C	.08	−.03	−.24	.11	.05	.19	−.03	0.81	[−.16, .09]
O	−.15	.03	−.05	.16	.05	.16	.06	−.03	0.65

Note. Values on the diagonal are variances, values on the lower triangular matrix are correlations, and values on the upper triangular matrix are 95 % posterior intervals of correlations.

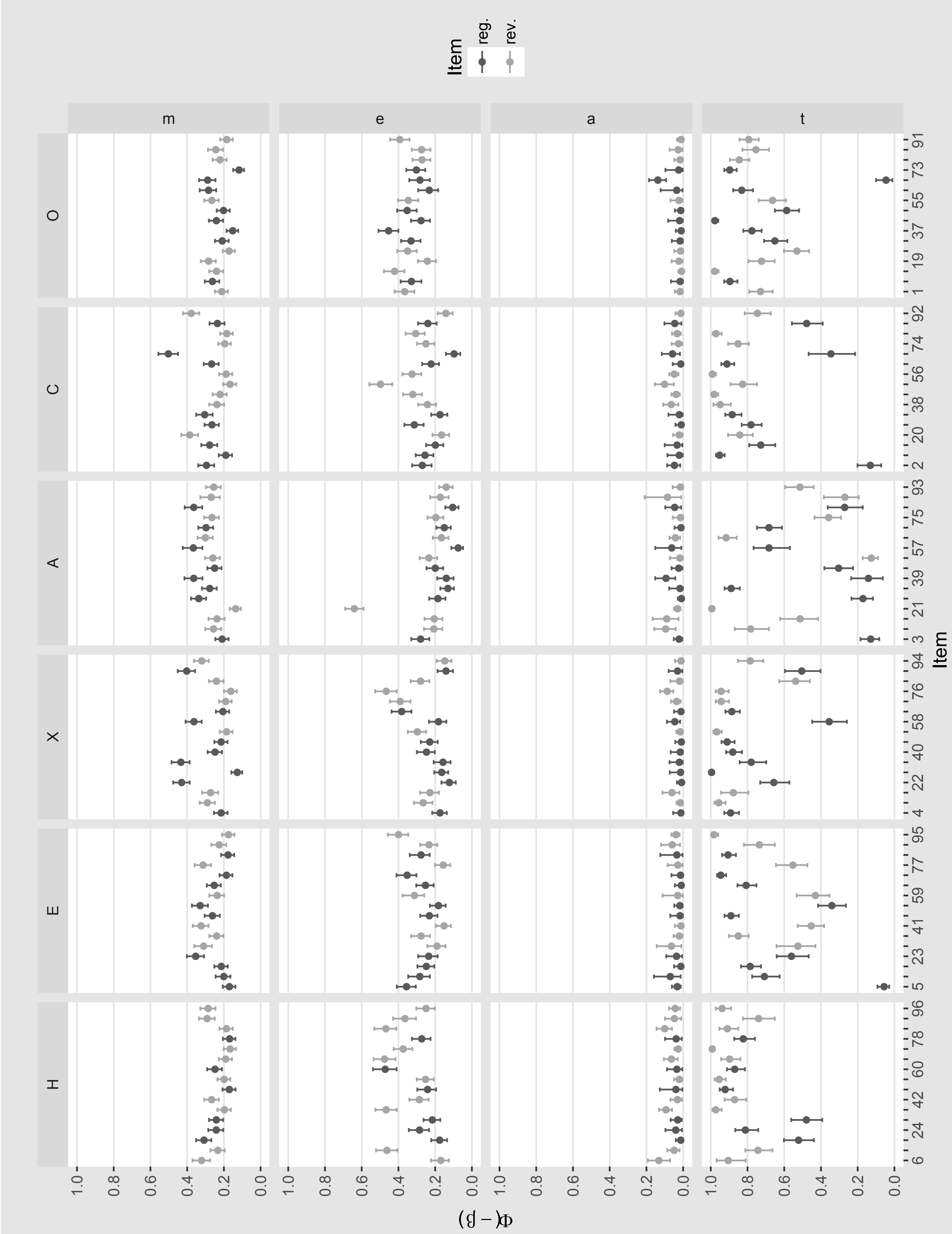

Figure 5. Posterior median and 95 % posterior intervals for the transformed item parameters as a function of (horizontally) the psychological process, (vertically) the corresponding HEXACO scale (Honesty-humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness, and Openness), and (black vs. gray) item coding direction.

relationship between ARS and the HEXACO traits is sparse, these correlations seem plausible at face value: Pretentious, hypocritical as well as sloppy, negligent persons (Ashton & Lee, 2007) were more prone to ARS compared to sincere as well as careful persons.

Apart from that, the intercorrelations of the six target traits were in the range from −.23 to .35 (see Table 1). Importantly, these correlations hardly differed from those estimated from a standard ordinal model, namely, a steps model[7] for six correlated target traits. The mean of the absolute differences between these two models was 0.019 and the largest difference was 0.046. This finding has two implications. First, controlling for response styles did not change the substantive conclusions about the intercorrelations in our data, which is in line with previous work (e.g., Hansjörg Plieninger, 2017). Second, the model entails, figuratively speaking, a crude dichotomization of the items into agreement versus disagreement. This, however, did not remove much substantive variance from the data at least concerning the intercorrelations.

When comparing the parameter estimates of the multidimensional Acquiescence Model with those from the single-domain version for honesty-humility above, the item parameters were almost identical with $r > .99$. While a similarly high correlation was observed for the person parameters θ_t measuring honesty-humility ($r = .98$), the response-style person-parameters showed smaller, but still high correlations between the two model version (.48 for MRS and .70 for both ERS and ARS). In line with our simulation studies, this illustrates that person parameters for response styles are estimated more precisely when using more items (96 vs. 16). In addition, the multidimensional model in principle benefits from the fact that a questionnaire with multiple content domains allows for a better discrimination between target traits and response styles.

A less restrictive Acquiescence Model. So far, the Acquiescence Model entailed the restriction that the ERS difficulties β_{e^*} in the ARS branch were set equal across all items as described above. Herein, we relaxed this assumption and estimated a β_{e^*j} parameter for every item. The posterior predictive checks, as described in Appendix B, revealed that this slightly improved model fit with respect to some but not all measures.

The item parameters correlated > .99 except for the ARS parameters β_a that showed a correlation of .48. The person parameters correlated > .99 except for the ARS parameters θ_a that showed a correlation of .84. However, the additional item parameters β_{e^*j} were estimated with small precision: Their intervals exhibited on average a width of 3.19 on the probit scale compared to widths of 0.30 for MRS and 1.19 for ARS, for example. Since these β_{e^*j} are defined conditionally on the a_{ij} parameter that had rather low probabilities as reported above, there was not much information provided by the data to estimate the β_{e^*j} parameters precisely. Thus, we chose the more restrictive model with more precise estimates rather than the model that provided slightly better fit.

Mixture vs. shift model. Above, we theoretically compared the Acquiescence Model, which defines ARS as a mixture process, with alternative accounts that model ARS as a shift process. To compare both accounts empirically, we additionally fit a classical factor-analysis model (e.g., Billiet & McClendon, 2000) to the data and retrieved the resulting person parameters θ^*_{ai} for the ARS factor. Additionally, we computed manifest proxy variables, which have been proposed as indices for acquiescence, for every person, namely, $A1$, the mean across all items before recoding (e.g., Couch & Keniston, 1960) and $A2$, the number of responses in the affirmative response categories *agree* and *strongly agree* (e.g., Billiet & McClendon, 2000).

Table 2 shows the correlations of these four measures, indicating that the shift model was in close agreement with the mean-index $A1$, much closer than with the count-index $A2$. The opposite pattern was found for the mixture model, which was more strongly related to $A2$ than to $A1$. These correlations thereby reflect the definition of the two indices $A1$ and $A2$, which are descriptively defined as shifts in item means versus changes in the number of affirmative responses, respectively. Concerning the model-based ARS estimates, the medium-sized correlation between the Acquiescence Model and the factor model (i.e., between θ_a and θ^*_a) indicated that there was substantial, but imperfect overlap between the definition of acquiescence in

[7]The steps model (Tutz, 1990; Verhelst et al., 1997) is an ordinal IRT model without response styles and was proposed as an alternative to, for example, the partial credit or the graded response model. It serves as natural comparison model herein, because it is also based on a tree-structure (De Boeck & Partchev, 2012).

Table 2
Correlations of Model-Based and Descriptive Measures of Acquiescence

	Model-based		Descriptive	
	θ_a	θ_a^*	$A1$	$A2$
θ_a	—	.36	.45	.71
θ_a^*		—	.96	.66
$A1$			—	.69

Note. θ_a: Acquiescence Model; θ_a^*: confirmatory factor model; $A1$: mean across all items; $A2$: number of agreements.

the mixture and the shift model. These results highlight that both models measure distinct albeit related constructs. Moreover, this implies that the definition of acquiescence in terms of a mixture or a shift process has consequences with respect to the measurement of ARS.

The fitted factor model and the Acquiescence Model differed not only with respect to the notion of shift vs. mixture but also in many other respect. Thus, we developed also a shift model within the IR-tree framework and defined a multidimensional parameter t_{ij} depending on both the target trait as well as acquiescence. The model and the comparison with the Acquiescence Model is described in Appendix A. The results mirrored those with the factor model even though the correlation of the ARS person parameters showed a higher correlation of .75 between the tree-shift and mixture model.

Discussion

We developed and tested a new model of acquiescence, a response style characterized by a preference for affirmative response categories. Inspired by MPT models, a popular model class in cognitive psychology, the new model builds on the work of Böckenholt (2012a) and explicitly assumes a theoretically motivated, tree-like structure of latent cognitive processes. As an extension of the original model, the new Acquiescence Model allows to capture not only ERS and MRS, but also ARS. All of these processes are modeled using an IRT approach by reparameterizing the probabilities of entering different states by additive person and item effects on the probit scale. Within the proposed model, agreement to an item is conceptualized as a mixture process and can either be due to a high target-trait level or due to ARS.

The Proposed Acquiescence Model

The original Böckenholt Model in Figure 1 assumes three qualitatively different processes. Whereas MRS directly leads to midpoint responses, the target trait leads to agreement with regular items (and disagreement with reversed items) conditional on non-MRS, in which case ERS determines whether more or less extreme responses are given. The Böckenholt Model of response styles and IR-tree models in general are characterized by these definitions of conditional response processes similar as in MPT models that are used in cognitive psychology (e.g., Erdfelder et al., 2009; Matzke et al., 2015). In psychometrics, this approach has been proven useful in both methodological and applied work (e.g., Böckenholt, 2017; Böckenholt & Meiser, 2017; Jeon & De Boeck, 2016; Khorramdel & von Davier, 2014; Hansjörg Plieninger & Meiser, 2014; Thissen-Roe & Thissen, 2013; Zettler, Lang, Hülsheger, & Hilbig, 2016). However, ARS—which is often seen as an especially important response style (e.g., Hansjörg Plieninger, 2017; Rammstedt, Goldberg, & Borg, 2010)—was not included in the Böckenholt Model, a disadvantage compared to alternative response style models (e.g., Johnson & Bolt, 2010). As a remedy, we proposed an extension of the Böckenholt Model, namely, the Acquiescence Model, which assumes an additional ARS process that leads to agreement with an item irrespective of its coding direction or content. Thus, the new model allows to disentangle four different processes—three response styles as well as the target trait.

Each of the four processes is comprised of a person parameter θ, which captures individual differences (e.g., in ARS-responding), and an item parameter β, which captures item specific effects. With respect to the target trait t_{ij}, these item effects correspond to item difficulties. With respect to response styles, these item effects capture difference between items in eliciting a particular response style. It is important to note that those style-related difficulties are not an indication of response styles being dependent on item *content*. Rather, "external" item features such as length, position, or complexity are possible explanations for this heterogeneity (e.g., De Jong et al., 2008), and Figure 5 illustrates that empirical data indeed exhibit such heterogeneity. While previous research mainly focused on person-level covariates of response styles (see Van Vaerenbergh & Thomas, 2013), future research may also focus on item-level covariates based on the item parameter estimates of the Acquiescence Model.

Future work may adapt the Acquiescence Model to items with more or less than five categories (see Böckenholt & Meiser, 2017; Hansjörg Plieninger & Meiser, 2014). Moreover, the difference between *agree* and *strongly agree* responses conditional on acquiescence may be of further interest. With respect to this "(ARS-)conditional ERS process", we opted for a parsimonious model with only a single item parameter β_{e^*}, even though, in the empirical illustration, the model with freely estimated β_{e^*j} parameters fitted slightly better according to some measures. Furthermore, we modeled response styles as stable across content domains, whereas the trait parameters t_{ij} were of course domain specific. Even though this is in line with previous work (e.g., Danner et al., 2015; Khorramdel & von Davier, 2014; Weijters et al., 2010a), other scholars have argued for a different approach (e.g., Ferrando et al., 2004), which might also be applied to the Acquiescence Model.

Comparison With Shift Models

Existing factor models of ARS (e.g., Billiet & McClendon, 2000; Ferrando et al., 2016) have the advantage that they are easy to interpret and often easy to implement in respective software. However, these models are committed to a specific interpretation of ARS in terms of an additive shift towards affirmative responses that can be compensated by low levels on the content trait. Often, however, ARS is defined as a response style that results in a higher proportion of *agree* responses. This definition is at odds with the factor model and better represented by the proposed Acquiescence Model, which assumes that affirmative responses emerge as a mixture of an ARS and a content-related process. In contrast, if acquiescence and disacquiescence are indeed opposite poles on a single dimension, then a shift approach as implemented in factor models is more appropriate. Note that a shift model of ARS can even be implemented within the IR-tree framework as shown in Appendix A. However, the important distinction between a shift and a mixture account of ARS has been largely overlooked in previous developments and applications. As a remedy, we developed a new model that allows to test the mixture account of ARS, thereby opening new directions for future research.

Even though the present paper mainly focused on the measurement of ARS, it is important to highlight the theoretical implications of any chosen measurement model. The literature reviewed in the Introduction showed several unresolved topics of acquiescence research, for example, the distinction between acquiescence, item-wording effects, and careless responding. Our discussion of shift and mixture approaches can help to shed new light on such questions, and the proposed models facilitates future research in this direction. Specifically, the theoretical motivation and implications of the different approaches discussed in the Introduction require further scrutiny. For instance, future research may investigate the relationship between the proposed mixture model and the dual-process theory of acquiescence presented by Knowles and Condon (1999). According to this theory, comprehension of an item or a belief in general requires initial acceptance of its content. Only in a second processing stage, the item's content is assessed more thoroughly, which may result in continued agreement or "un-accepting" (see Gilbert, 1991; Mandelbaum, 2014). Since acquiescence is the default processing mode, premature output after the first stage results in agreement and thus serves as a theoretical explanation of ARS (Knowles & Condon, 1999).

Hierarchical MPT Models and Bayesian Estimation

The proposed model belongs to the general class of hierarchical MPT models (Heck, Arnold, & Arnold, 2018; Klauer, 2010; Matzke et al., 2015), which subsume psychometric IR-tree models as special cases (Böckenholt, 2012a; De Boeck & Partchev, 2012; Tutz, 1990). We believe that hierarchical MPT models provide a general and fruitful framework for future developments and applications based on the idea and principles of "cognitive psychometrics" (Riefer et al., 2002). For example, hierarchical MPT models might shed new light on the 3PL model and facilitate modeling of guessing behavior more generally (see also von Davier, 2009). The 3PL model assumes that, in an ability test, both correct guesses and valid knowledge can lead to correct responses. Hence, the model assumes a mixture of two processes similar to ARS and the target trait in the Acquiescence Model, and can thus be interpreted as a hierarchical MPT model.

Concerning parameter estimation, the recoding procedure previously used to obtain maximum-likelihood estimates with software for multidimensional IRT models (Böckenholt, 2012a; De Boeck & Partchev, 2012; Jeon & De Boeck, 2016) was no longer applicable in the Acquiescence Model. This is due to the mixture structure on the level of item-person combinations, according to which the response probability of an affirmative response is defined as the sum of two branch probabilities. As a remedy, we adapted a Bayesian implementation of hierarchical MPT models (Klauer, 2010; Matzke et al., 2015). Besides providing virtually identical estimates as the ML procedure in case of the Böckenholt Model, this gives the researcher great flexibility to fit complex models such as the 9-dimensional Acquiescence Model for the HEXACO data presented above. In two simulation studies, we showed that parameter recovery of the Acquiescence Model was satisfactory. Moreover, fitting the model to data generated without acquiescence (i.e., the standard Böckenholt Model) did not affect conclusions substantially because (a) DIC was likely to select the correct model, and (b) the Acquiescence Model empirically reduced to the Böckenholt Model—that is, the ARS item parameters β_a became extremely large and the ARS person variance $\sigma^2_{\theta_a}$ became relatively small (thereby implying a very low probability of ARS responses). Taken together, these results show that the proposed Acquiescence Model provides a useful generalization of the Böckenholt Model.

Empirical Illustration

The empirical example illustrated that the prevalence of acquiescence was rather low for the German version of the HEXACO-PI-R in the current sample. On the one hand, this finding is reassuring from an assessment perspective, because higher ARS levels might raise validity concerns. On the other hand, the low prevalence of acquiescence makes a precise estimation of the ARS parameters difficult, a limitation that was counteracted by using a large number of 96 items. Note that the detrimental effects of acquiescence can be relatively small (e.g., Hansjörg Plieninger, 2017; Rorer, 1965; Savalei & Falk, 2014). However, research on ARS is not only important to correct possible biases in target-trait estimates, but also to fully understand and explain the latent processes that underly responses to questionnaire items. Based on a solid understanding of response styles, it is possible to develop appropriate models and to design questionnaires, tests, and testing situations such that test scores have a high reliability and validity.

Conclusion

In sum, we proposed the Acquiescence Model to generalize and improve an already successful response style model (Böckenholt, 2012a). Thereby, we provide an answer to the question how to account for the empirically relevant and theoretically interesting phenomenon of ARS within the psychologically meaningful tree-like structure of IR-tree models. By modeling agreement as a mixture of acquiescence and target trait, we shed light on the question how to define and measure acquiescent response behavior. To address such theoretical questions in general, we advocate the use of hierarchical MPT models that explicitly account for latent response processes and thereby provide a powerful framework at the interface of psychometrics and cognitive psychology.

References

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*, 150–166. doi:10.1177/1088868306294907

Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual differences in response scale use: Mixed Rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences*, *40*, 1235–1245. doi:10.1016/j.paid.2005.10.018

Babcock, B. (2011). Estimating a noncompensatory IRT Model using Metropolis within Gibbs sampling. *Applied Psychological Measurement*, *35*, 317–329. doi:10.1177/0146621610392366

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*, 57–86. doi:10.3758/BF03210812

Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*, 143–156. doi:10.1509/jmkr.38.2.143.18840

Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin*, *76*, 186–204. doi:10.1037/h0031474

Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods & Research*, *36*, 542–562. doi:10.1177/0049124107313901

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, *7*, 608–628. doi:10.1207/S15328007SEM0704_5

Böckenholt, U. (2012a). Modeling multiple response processes in judgment and choice. *Psychological Methods*, *17*, 665–678. doi:10.1037/a0028111

Böckenholt, U. (2012b). The cognitive-miser response model: Testing for intuitive and deliberate reasoning. *Psychometrika*, *77*, 388–399. doi:10.1007/s11336-012-9251-y

Böckenholt, U. (2014). Modeling motivated misreports to sensitive survey questions. *Psychometrika*, *79*, 515–537. doi:10.1007/s11336-013-9390-9

Böckenholt, U. (2017). Measuring response styles in Likert items. *Psychological Methods*, *22*, 69–83. doi:10.1037/met0000106

Böckenholt, U., & Meiser, T. (2017). Response style analysis with threshold and multi-process IRT models: A review and tutorial. *British Journal of Mathematical and Statistical Psychology*, *70*, 159–181. doi:10.1111/bmsp.12086

Bolt, D. M., & Lall, V. F. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte Carlo. *Applied Psychological Measurement*, *27*, 395–414. doi:10.1177/0146621603258350

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. doi:10.18637/jss.v076.i01

Couch, A., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *The Journal of Abnormal and Social Psychology*, *60*, 151–174. doi:10.1037/h0040372

Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, *6*, 475–494. doi:10.1177/001316444600600405

Danner, D., Aichholzer, J., & Rammstedt, B. (2015). Acquiescence in personality questionnaires: Relevance, domain specificity, and stability. *Journal of Research in Personality*, *57*, 119–130. doi:10.1016/j.jrp.2015.05.004

De Boeck, P., & Partchev, I. (2012). IRTrees: Tree-based item response models of the GLMM family. *Journal of Statistical Software*, *48*(1), 1–28. doi:10.18637/jss.v048.c01

De Jong, M. G., Steenkamp, J.-B. E. M., Fox, J.-P., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research*, *45*, 104–115. doi:10.1509/jmkr.45.1.104

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*, 1–38. Retrieved from http://www.jstor.org/stable/2984875

Denwood, M. J. (2016). runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software*, *71*(1), 1–25. doi:10.18637/jss.v071.i09

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models. *Zeitschrift für Psychologie/Journal of Psychology*, *217*, 108–124. doi:10.1027/0044-3409.217.3.108

Falk, C. F., & Cai, L. (2016). A flexible full-information approach to the modeling of response styles. *Psychological Methods*, *21*, 328–347. doi:10.1037/met0000059

Ferrando, P. J., Condon, L., & Chico, E. (2004). The convergent validity of acquiescence: An empirical study relating balanced scales and separate acquiescence scales. *Personality and Individual Differences*, *37*, 1331–1340. doi:10.1016/j.paid.2004.01.003

Ferrando, P. J., Morales-Vives, F., & Lorenzo-Seva, U. (2016). Assessing and controlling acquiescent responding when acquiescence and content are related: A comprehensive factor-analytic approach. *Structural Equation Modeling*, *23*, 713–725. doi:10.1080/10705511.2016.1185723

Fox, J.-P. (2010). *Bayesian item response modeling: Theory and applications*. doi:10.1007/978-1-4419-0742-4

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: CRC.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, *46*, 107–119. doi:10.1037/0003-066X.46.2.107

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*, 264–284. doi:10.3758/s13428-017-0869-7

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1593–1623. Retrieved from http://jmlr.org/

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, *59*, 21–47. doi:10.1007/bf02294263

Hütter, M., & Klauer, K. C. (2016). Applying processing trees in social psychology. *European Review of Social Psychology*, *27*, 116–159. doi:10.1080/10463283.2016.1212966

Jeon, M., & De Boeck, P. (2016). A generalized item response tree model for psychological assessments. *Behavior Research Methods*, *48*, 1070–1085. doi:10.3758/s13428-015-0631-y

Johnson, T. R., & Bolt, D. M. (2010). On the use of factor-analytic multinomial logit item response models to account for individual differences in response style. *Journal of Educational and Behavioral Statistics*, *35*, 92–114. doi:10.3102/1076998609340529

Johnson, T. R., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology*, *36*, 264–277. doi:10.1177/0022022104272905

Kam, C. C. S., & Zhou, M. (2015). Does acquiescence affect individual items consistently? *Educational and Psychological Measurement*, *75*, 764–784. doi:10.1177/0013164414560817

Khorramdel, L., & von Davier, M. (2014). Measuring response styles across the Big Five: A multiscale extension of an approach using multinomial processing trees. *Multivariate Behavioral Research*, *49*, 161–177. doi:10.1080/00273171.2013.866536

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*, 70–98. doi:10.1007/s11336-009-9141-0

Knowles, E. S., & Condon, C. A. (1999). Why people say "yes": A dual-process theory of acquiescence. *Journal of Personality and Social Psychology*, *77*, 379–386. doi:10.1037/0022-3514.77.2.379

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, *50*, 537–567. doi:10.1146/annurev.psych.50.1.537

Lee, K., & Ashton, M. C. (2016). Psychometric properties of the HEXACO-100. *Assessment*. Advance online publication. doi:10.1177/1073191116659134

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. New York, NY: Cambridge University.

Lentz, T. F. (1938). Acquiescence as a factor in the measurement of personality [Abstract]. *Psychological Bulletin*, *35*, 689. doi:10.1037/h0055433

Levy, R. (2011). Posterior predictive model checking for conjunctive multidimensionality in item response theory. *Journal of Educational and Behavioral Statistics*, *36*, 672–694. doi:10.3102/1076998611410213

Li, T., Xie, C., & Jiao, H. (2017). Assessing fit of alternative unidimensional polytomous IRT models using posterior predictive model checking. *Psychological Methods*, *22*, 397–408. doi:10.1037/met0000082

Mandelbaum, E. (2014). Thinking is believing. *Inquiry*, *57*, 55–96. doi:10.1080/0020174X.2014.858417

Marsh, H. W., Scalas, L. F., & Nagengast, B. (2010). Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem Scale: Traits, ephemeral artifacts, and stable response styles. *Psychological Assessment*, *22*, 366–381. doi:10.1037/a0019225

Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E.-J. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, *80*, 205–235. doi:10.1007/s11336-013-9374-9

Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods*, *11*, 344–362. doi:10.1037/1082-989X.11.4.344

Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, *44*, 1539–1550. doi:10.1016/j.paid.2008.01.010

Mirowsky, J., & Ross, C. E. (1991). Eliminating defense and agreement bias from measures of the sense of control: A 2 × 2 index. *Social Psychology Quarterly*, *54*, 127–145. doi:10.2307/2786931

Moshagen, M., Hilbig, B. E., & Zettler, I. (2014). Faktorenstruktur, psychometrische Eigenschaften und Messinvarianz der deutschsprachigen Version des 60-Item HEXACO Persönlichkeitsinventars [Factor structure, psychometric properties, and measurement invariance of the German-language version of the 60-item HEXACO personality inventory]. *Diagnostica*, *60*, 86–97. doi:10.1026/0012-1924/a000112

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (Vol. 1, pp. 17–59). San Diego, CA: Academic Press.

Plieninger, H. [Hansjörg]. (2017). Mountain or molehill? A simulation study on the impact of response styles. *Educational and Psychological Measurement*, *77*, 32–53. doi:10.1177/0013164416636655

Plieninger, H. [Hansjörg]. (2018). *Towards a deeper understanding of response styles through psychometrics* (Doctoral dissertation, University of Mannheim, Mannheim, Germany). Retrieved from https://ub-madoc.bib.uni-mannheim.de/44325/

Plieninger, H. [Hansjörg], & Meiser, T. (2014). Validity of multiprocess IRT models for separating content and response styles. *Educational and Psychological Measurement*, *74*, 875–899. doi:10.1177/0013164413514998

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2013)*, Vienna, Austria. Retrieved from https://www.r-project.org/conferences/DSC-2003/Proceedings

R Core Team. (2016). R: A language and environment for statistical computing. Retrieved from https://www.r-project.org

Rammstedt, B., Goldberg, L. R., & Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality*, *44*, 53–61. doi:10.1016/j.jrp.2009.10.005

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*, 318–339. doi:10.1037/0033-295X.95.3.318

Riefer, D. M., Knapp, B. R., Batchelder, W. H., Bamber, D., & Manifold, V. (2002). Cognitive psychometrics: Assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment*, *14*, 184–201. doi:10.1037/1040-3590.14.2.184

Rorer, L. G. (1965). The great response-style myth. *Psychological Bulletin*, *63*, 129–156. doi:10.1037/h0021888

Rost, J., Carstensen, C. H., & von Davier, M. (1997). Applying the mixed Rasch model to personality questionnaires. In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 324–332). Münster, Germany: Waxmann.

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, *12*, 573–604. doi:10.3758/BF03196750

Savalei, V., & Falk, C. F. (2014). Recovering substantive factor loadings in the presence of acquiescence bias: A comparison of three approaches. *Multivariate Behavioral Research*, *49*, 407–424. doi:10.1080/00273171.2014.931800

Schmittmann, V. D., Dolan, C. V., Raijmakers, M. E. J., & Batchelder, W. H. (2010). Parameter identification in multinomial processing tree models. *Behavior Research Methods*, *42*, 836–846. doi:10.3758/BRM.42.3.836

Shulruf, B., Hattie, J., & Dixon, R. (2008). Factors affecting responses to Likert type questionnaires: Introduction of the ImpExp, a new comprehensive model. *Social Psychology of Education*, *11*, 59–78. doi:10.1007/s11218-007-9035-x

Sinharay, S., Johnson, M. S., & Stern, H. S. (2006). Posterior predictive assessment of item response theory models. *Applied Psychological Measurement*, *30*, 298–321. doi:10.1177/0146621605285517

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology*, *94*, 718–737. doi:10.1037/0022-3514.94.4.718

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *64*, 583–639. doi:10.1111/1467-9868.00353

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research*, *45*, 116–131. doi:10.1509/jmkr.45.1.116

ten Berge, J. M. F. (1999). A legitimate case of component analysis of ipsative measures, and partialling the mean as an alternative to ipsatization. *Multivariate Behavioral Research*, *34*, 89–102. doi:10.1207/s15327906mbr3401_4

Thissen-Roe, A., & Thissen, D. (2013). A two-decision model for responses to Likert-type items. *Journal of Educational and Behavioral Statistics*, *38*, 522–547. doi:10.3102/1076998613481500

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, *103*, 299–314. doi:10.1037/0033-2909.103.3.299

Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology*, *43*, 39–55. doi:10.1111/j.2044-8317.1990.tb00925.x

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*, 195–217. doi:10.1093/ijpor/eds021

Verhelst, N. D., Glas, C. A. W., & de Vries, H. H. (1997). A steps model to analyze partial credit. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 123–138). doi:10.1007/978-1-4757-2691-6_7

von Davier, M. (2009). Is there need for the 3PL model? Guess what? *Measurement*, *7*, 110–114. doi:10.1080/15366360903117079

Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research*, *49*, 737–747. doi:10.1509/jmr.11.0368

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods*, *18*, 320–334. doi:10.1037/a0032121

Weijters, B., Geuens, M., & Schillewaert, N. (2010a). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement*, *34*, 105–121. doi:10.1177/0146621609338593

Weijters, B., Geuens, M., & Schillewaert, N. (2010b). The stability of individual response styles. *Psychological Methods*, *15*, 96–110. doi:10.1037/a0018721

Wetzel, E., Böhnke, J. R., & Brown, A. (2016). Response biases. In F. T. L. Leong, D. Bartram, F. Cheung, K. F. Geisinger, & D. Iliescu (Eds.), *The ITC International Handbook of Testing and Assessment* (pp. 349–363). doi:10.1093/med:psych/9780199356942.003.0024

Wetzel, E., & Carstensen, C. H. (2017). Multidimensional modeling of traits and response styles. *European Journal of Psychological Assessment*, *33*, 352–364. doi:10.1027/1015-5759/a000291

Winkler, J. D., Kanouse, D. E., & Ware, J. (1982). Controlling for acquiescence response set in scale development. *Journal of Applied Psychology*, *67*, 555–561. doi:10.1037/0021-9010.67.5.555

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, *30*, 187–213. doi:10.1111/j.1745-3984.1993.tb00423.x

Zaller, J., & Feldman, S. (1992). A simple theory of the survey response: Answering questions versus revealing preferences. *American Journal of Political Science*, *36*, 579. Retrieved from http://www.jstor.org/stable/2111583

Zettler, I., Lang, J. W. B., Hülsheger, U. R., & Hilbig, B. E. (2016). Dissociating indifferent, directional, and extreme responding in personality data: Applying the three-process model to self- and observer reports. *Journal of Personality*, *84*, 461–472. doi:10.1111/jopy.12172

Zhu, X., & Stone, C. A. (2012). Bayesian comparison of alternative graded response models for performance assessment applications. *Educational and Psychological Measurement*, *72*, 774–799. doi:10.1177/0013164411434638

Appendix A
An IR-Tree Shift Model for Acquiescence

We contrasted our proposed mixture model for acquiescence with the most prominent alternative model, namely, the two-factor or random-intercept model (e.g., Billiet & McClendon, 2000; Maydeu-Olivares & Coffman, 2006; Mirowsky & Ross, 1991). We showed that such a two-factor model conceptualizes ARS in terms of a compensatory shift process. This approach is most often used in the context of factor analysis, but it may also be implemented in the context of an IRT model (e.g., Falk & Cai, 2016).

We deliberately chose the two-factor approach, because of its widespread use and in order to make the distinction between the models most clear. Even though the theoretical comparison of the model is enlightening, an empirical comparison is difficult to make, because the models differ in many respects. Furthermore, it is also possible to implement a shift model within the IR-tree framework, as alluded to by two anonymous reviewers.

This is accomplished by allowing for multidimensional nodes (see Jeon & De Boeck, 2016) in the Böckenholt Model (depicted in Figure 1). More precisely, we reparameterize t_{ij} as follows:

$$t_{ij} = \Phi(\theta_{ti} + (-1)^r \theta_{a'i} - \beta_{tj})$$
$$\text{with } r = \begin{cases} 0, & \text{for regular items} \\ 1, & \text{for reversed items.} \end{cases} \tag{24}$$

That is, agreement is, for regular items, most likely for persons high on the target trait θ_{ti} and/or high on ARS $\theta_{a'i}$. For reversed items, in contrast, agreement is most likely for persons *low* on the target trait θ_{ti} and/or *high* on ARS $\theta_{a'i}$.

This model has the following implications: First, it is a shift model since ARS ($\theta_{a'i}$) is added or subtracted, on the latent scale, from the target trait (θ_{ti}). Note, however, that this shift process does not involve the middle category (in contrast to two-factor models as discussed above). Nevertheless, negative values of $\theta_{a'i}$ imply disacquiescence as in the two-factor model. Second, the model is a compensatory model, because target trait and ARS may outweigh each other. Third, the model does not contain ARS-specific item parameters even though ARS may have an effect on β_{tj}. Thus, the model is more parsimonious/restrictive compared to the Acquiescence Model. In summary, the tree-shift model is somewhere in between the two-factor model and the Acquiescence Model and shares features with both of them.

We also fit the tree-shift model to the HEXACO data, and DIC equaled 122,684: This is superior to the Böckenholt Model indicating again the importance of taking ARS into account; but fit is worse compared to the Acquiescence Model highlighting once more that the proposed mixture model is a reasonable alternative to the shift account. When comparing the tree-shift model and the Acquiescence Model, the item parameters showed strong correlations of .99 (MRS), .99 (ERS), and .96 (target traits). The same holds for the person parameters which correlated between .98 and .99 for the six target traits, 1.00 for MRS and ERS, and .63 for ARS. Similar to the comparison with the two-factor model reported in Table 2, this shows that ARS in the mixture and ARS in the shift approach are highly similar albeit distinct constructs.

We also investigated the relative fit of the tree-shift and the Acquiescence Model via means of posterior predictive checks as reported in Appendix B. Therein, the Acquiescence Model outperformed the tree-shift model. However, the comparison is based only on a single, illustrative data set, and the aim of the present work was not to rule out the shift approach. Rather, the proposed model allows to shed new light on the old phenomenon of acquiescence.

Appendix B
Posterior Predictive Checks

Model fit was further evaluated by means of posterior predictive checks (e.g., Gelman et al., 2014; M. D. Lee & Wagenmakers, 2013; Shulruf et al., 2008), that is, by assessing the discrepancy between observed data and data predicted on the basis of the posterior samples. All reported analyses are based on a subset of 1,000 iterations.

Graphical Model Checks

For every item, a response distribution predicted by the model was obtained in each iteration. Aggregated across iterations, we compared the 68 % and 95 % posterior intervals against the observed response distribution to test whether the model accurately accounted for the observed data. This was done separately for the Acquiescence Model, the Böckenholt Model, and a steps model that does not account for response styles at all.

Posterior predictive checks are illustrated in Figure B1A, namely, for respondents above the 90th percentile of the ARS distribution and for the three items most susceptible to ARS (i.e., extreme $\hat{\beta}_{aj}$). If acquiescence affected response behavior, the Acquiescence Model should outperform the two other models because it explicitly accounts for ARS. Figure B1A shows that this was indeed the case. The Acquiescence Model was superior to the two competitors especially in predicting response frequencies in Category 4 (i.e., *agree*). Note that this pattern was observed for all 96 items: The coverage rate (i.e., the proportion of 95 % posterior intervals covering the observed frequencies) was 95 % for the Acquiescence Model, 90 % for the Böckenholt Model, and only 83 % for the Steps Model.

Graphical Cross Validation

We also performed a cross-validation check to assess whether the flexibility of the Acquiescence Model led to overfitting and poor predictions of new data. Therefore, the quality of out-of-sample predictions was investigated using the remaining 512 respondents from the sample of Moshagen et al. (2014) as a cross-validation data set. The bars in Figure B1B show the observed frequencies for these respondents for the same three items as before, whereas the error bars indicate the predicted intervals for 512 hypothetical persons sampled from the multivariate posterior distribution of the person parameters. The Acquiescence Model predicted the new, empirical response frequencies very well with all 95 % and most 68 % posterior intervals covering the observed data. Whereas the steps model performed comparatively well, the Böckenholt Model provided a worse prediction, especially for the *strongly agree*-category. With respect to all 96 items, the Acquiescence Model with a coverage rate of 92 % again outperformed the Böckenholt Model with 87 % coverage, while the steps model performed slightly better with 97 % coverage. In summary, the posterior predictive checks indicated that the Acquiescence Model successfully predicted the fitted data and also new response frequencies thereby corroborating the model-selection results based on DIC above.

Posterior Predictive *p*-Values

Furthermore, model fit was assessed by means of posterior predictive *p*-values (PPP-values). These are based on different discrepancy measures that quantify the deviation between observations and predictions. The selected discrepancy measures should be sensitive to deviations relevant to the application at hand (Gelman et al., 2014; Sinharay, Johnson, & Stern, 2006). Based on the literature, we selected three measures: The *item score distribution* is based on the classical notion of residuals and measures the discrepancy between observed frequencies and posterior predictive frequencies for each item using Pearson's X^2 statistic (e.g., Zhu & Stone, 2012). Thus, this is a more formal way to quantify the analyses reported in Figure B1A. Furthermore, we used two other summary statistics that focus on the relationship between items, local dependence, and a questionnaire's dimensionality. First, Yen's Q_3 (Yen, 1993) measures the association between pairs of items after accounting for the latent variables (e.g., Levy, 2011; Li, Xie, & Jiao, 2017; Zhu & Stone, 2012). Since ARS may have an effect on relationships between items, this statistic should be able to discriminate between models with and without ARS. Second, the global odds ratio (*OR*) measures the association between pairs of items, and can be used with polytomous items after dichotomization (here, categories $1, 2, 3$ vs. $4, 5$; e.g., Li et al., 2017; Sinharay et al., 2006; Zhu & Stone, 2012). Because of this comparison of agreement vs. non-agreement, this statistic seemed well-suited to assess the fit of the models at hand.

Even though PPP-values are similarly defined as frequentist *p*-values via tail-area probabilities, they are not necessarily uniformly distributed under the null hypothesis and cannot always be used in the same way (e.g., Sinharay et al., 2006). Thus, they are interpreted herein more as a relative than an absolute measure of model fit. For the empirical example concerning the 96 HEXACO items, we compared the relative fit of five models, namely, the Böckenholt Model, the Acquiescence Model, an unrestricted[8] Acquiescence Model (u-Acquiescence), an ordinal steps model without re-

[8] As discussed on page 6 and 17, all 96 β_{e*j} parameters were constrained to be equal in the Acquiescence Model, whereas they were freely estimated in the unrestricted Acquiescence Model.

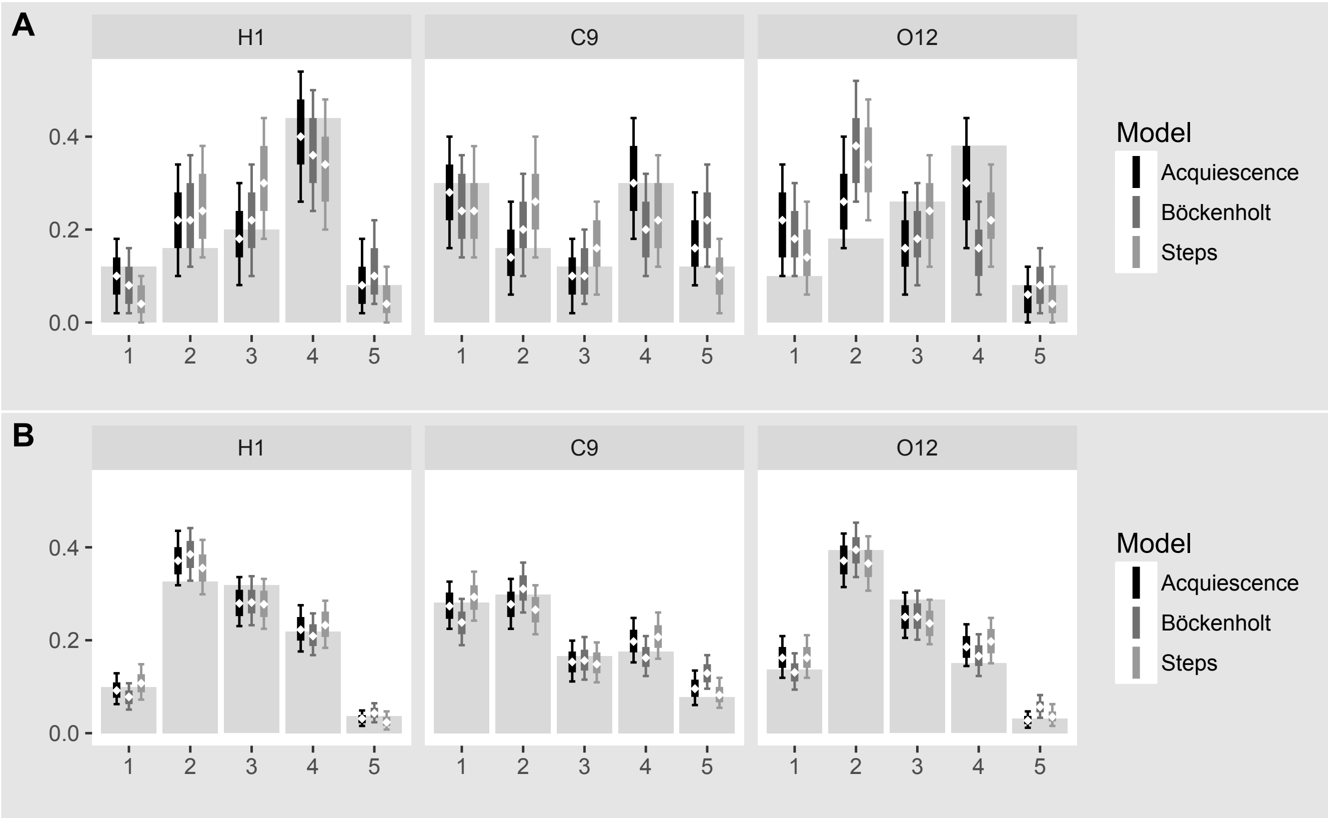

Figure B1. Posterior predictive checks for three selected items. The thin (fat) error bars represent 95 % (68 %) posterior intervals from the posterior-predictive distribution of three different models. Panel A displays the observed frequencies of the subset of persons from the original sample with ARS estimates above the 90th percentile. Panel B displays the observed frequencies of all persons in the cross-validation sample and the out-of-sample predictions for hypothetical respondents.

sponse styles, and a tree-shift model discussed in Appendix A. For the 96 items, we obtained 96 PPP-values for X^2, and $\binom{96}{2} = 4560$ PPP-values for Q_3 and OR. Figure B2 gives, for each model, the percentage of PPP-values that were below and above the conventional levels of 5 % and 95 %, respectively.

The steps model showed an excellent fit according to the item score distribution X^2 followed by the unrestricted and restricted Acquiescence Model; the Böckenholt and the tree-shift model showed impaired fit. Regarding Yen's Q_3 and OR, the Böckenholt Model and the steps model performed worse than the three models taking acquiescence into account highlighting again the importance of this response style. There were no substantive differences between the Acquiescence Model, its unrestricted version, and the shift model. Summarizing all three measures, it turned out that the proposed Acquiescence Model provided a relatively good fit and was only slightly outperformed by the unrestricted Acquiescence Model according to X^2.

Appendix C
Recovery of ARS Parameters

In an additional simulation study, we checked that the decreased recovery of the ARS person parameters in Study 1 was due to the low absolute prevalence of ARS (and not due to any insufficiencies of the model or estimation method). For this purpose, we simulated data for two conditions that differed in the prevalence of acquiescence. The first condition with 20 items was identical to the first condition in Study 1 (see Figure 4). In the second condition, a higher prevalence of ARS was realized: The true ARS item parameters β_a were drawn from a distribution identical to that for MRS and ERS (i.e., β_m and β_e), that is, with $\mu_{\beta_a} = \Phi^{-1}(.70)$. The results illustrated in Figure C1 showed that the impai-

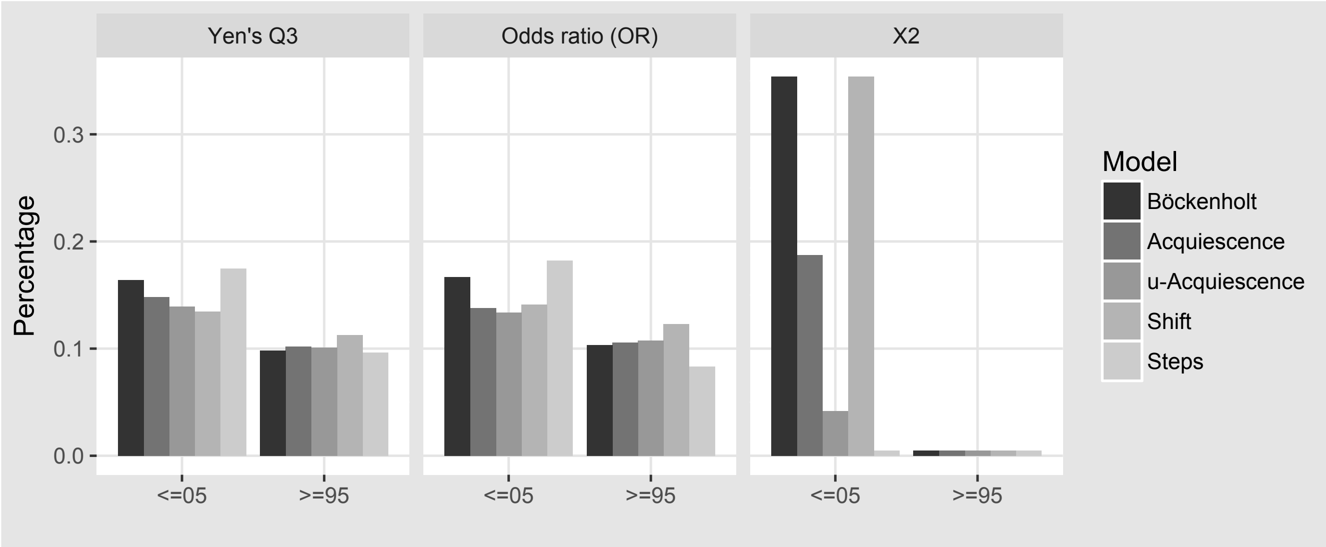

Figure B2. Posterior predictive *p*-values. Comparison of five models using the three discrepancy measures Yen's Q_3, odds ratio (*OR*), and item score distribution (X^2). Displayed are the percentages of PPP-values below and above conventional criteria of 5 % and 95 %.

red recovery of the ARS-parameters in Figure 4 was mainly due to the low, but realistic prevalence of ARS and not due to insufficiencies of the Acquiescence Model. In contrast, the precision of the target-trait parameter decreased only slightly, indicating the Acquiescence Model provides unbiased estimates even when ARS is highly prevalent. Overall, this simulation shows the trade-off in measuring the target trait versus ARS responding—a trade-off that is theoretically predicted by the mixture structure illustrated in Figure 2.

Appendix D
Graphical Convergence Diagnostic

When fitting the models both in the empirical study as well as in the simulation studies, careful attention was paid to convergence of the MCMC sampler. In this appendix, we focus on the model for all six HEXACO scales reported in the empirical study and showcase the convergence of four selected model parameters. For this illustration, we selected parameters that were of particular substantive interest and showed comparatively slow convergence. In particular, convergence is shown (a) for the lowest β_{aj}-parameter (i.e., the item most easily eliciting ARS responses, namely, $\beta_{a,67}$), (b) for the highest θ_{ai}-parameter (i.e., the person scoring highest on ARS, namely, $\theta_{a,294}$), (c) for the variance of the ARS person parameters $\sigma^2_{\theta_a}$ (i.e., $\mathbf{\Sigma}_{3,3}$), and (d) for the covariance between ARS and honesty-humility (i.e., $\mathbf{\Sigma}_{4,3}$). Note that both the θ- and the β-parameter are displayed on the original probit scale. For example, the estimate for $\beta_{a,67}$ on the probit scale is 1.09 [0.89, 1.32], which corresponds to .86 [.81, .91] on the probability scale, which was reported above (see also Figure 5). Note further, that the estimate for $\mathbf{\Sigma}_{3,3}$ is also reported in Table 1, but the figure below shows the covariance $\mathbf{\Sigma}_{4,3}$ whereas Table 1 contains the corresponding correlation.

In all four panels in Figure D1, convergence is indicated by the following features: (a) the densities resulting from the six different chains are almost identical, (b) the "point estimates" as represented by the running mean are almost identical across chains, (c) the traceplots show nicely mixing chains without any irregularities, and (d) almost no autocorrelation is observed (a showcase of the capabilities of Stan).

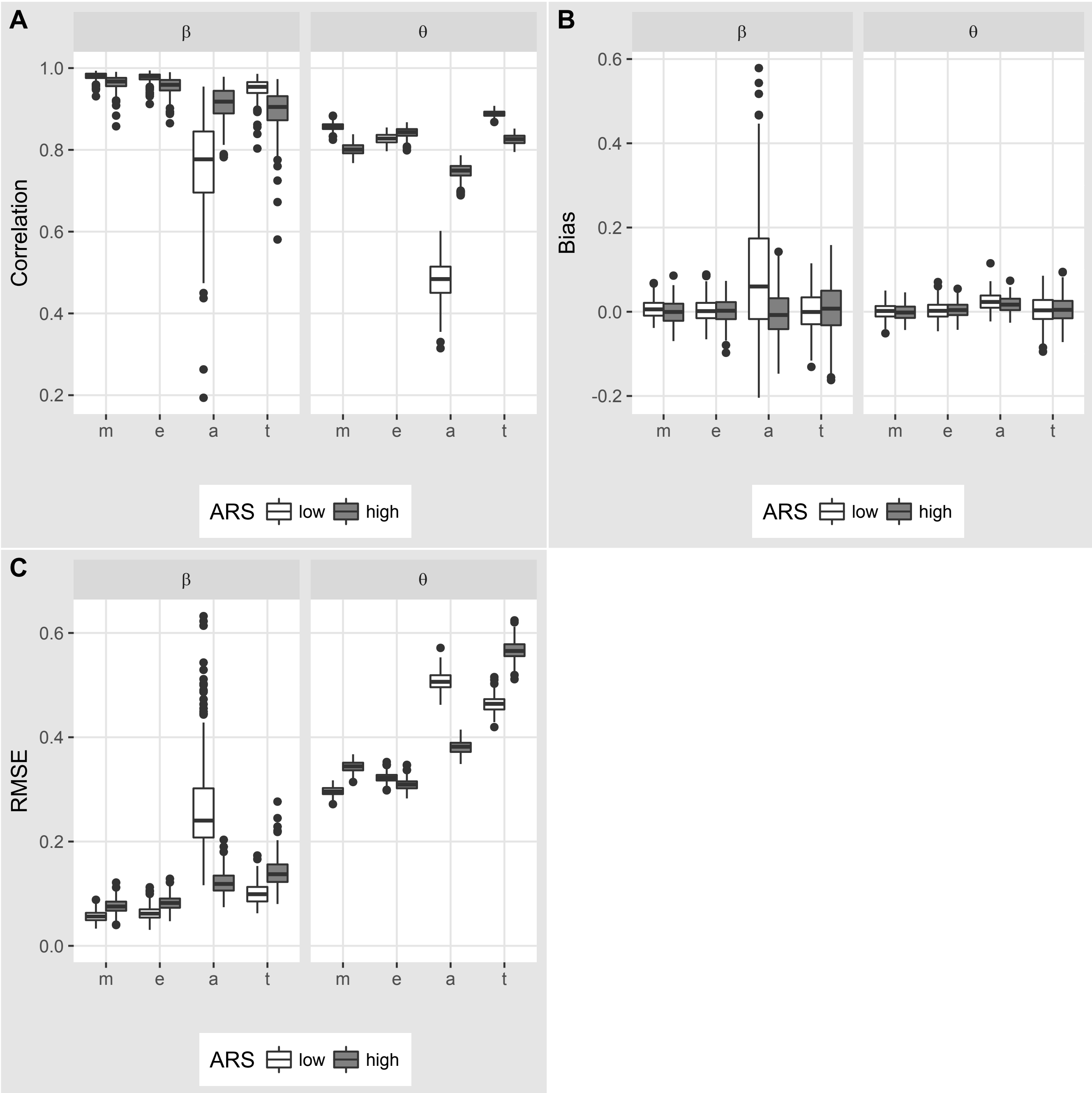

Figure C1. Boxplots illustrating parameter recovery of the item and person parameters of the Acquiescence Model when generating data with either a low (white) or high (gray) prevalence of acquiescence.

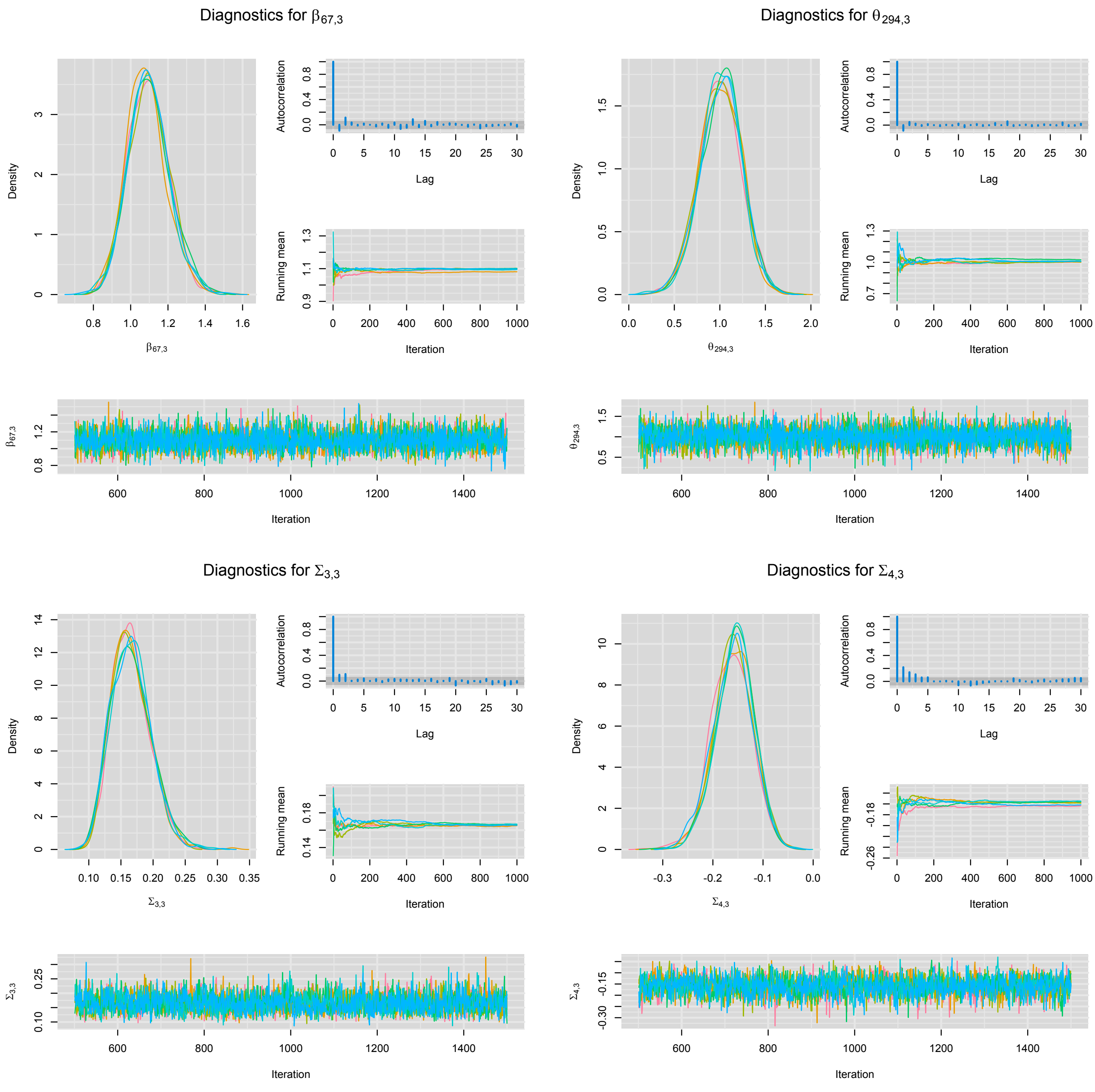

Figure D1. Graphical convergence diagnostics for four parameters. Displayed within each of the four panels are (in clockwise order starting from top left) a density plot, an autocorrelation plot, a plot of the running mean, and a traceplot (for each of six chains).