Received: 27 January 2021 | Accepted: 28 December 2021

DOI: 10.1111/rssa.12810

ORIGINAL ARTICLE

ROYAL STATISTICAL SOCIETY | Series A Statistics in Society

A multidimensional pairwise comparison model for heterogeneous perceptions with an application to modelling the perceived truthfulness of public statements on COVID-19

Qiushi Yu | **Kevin M. Quinn**

Department of Political Science, University of Michigan, Ann Arbor, USA

Correspondence
Qiushi Yu, Department of Political Science, University of Michigan, 505 South State Street, 5700 Haven Hall, Ann Arbor, MI 48109, USA.
Email: yuqiushi@umich.edu

Funding information
U.S. National Science Foundation, Grant/Award Number: SES 16-59922

Abstract

Pairwise comparison models are an important type of latent attribute measurement model with broad applications in the social and behavioural sciences. Current pairwise comparison models are typically unidimensional. The existing multidimensional pairwise comparison models tend to be difficult to interpret and they are unable to identify groups of raters that share the same rater-specific parameters. To fill this gap, we propose a new multidimensional pairwise comparison model with enhanced interpretability which explicitly models how object attributes on different dimensions are differentially perceived by raters. Moreover, we add a Dirichlet process prior on rater-specific parameters which allows us to flexibly cluster raters into groups with similar perceptual orientations. We conduct simulation studies to show that the new model is able to recover the true latent variable values from the observed binary choice data. We use the new model to analyse original survey data regarding the perceived truthfulness of statements on COVID-19 collected in the summer of 2020. By leveraging the strengths of the new model, we find that the partisanship of the speaker and the partisanship of the respondent account for the majority of

the variation in perceived truthfulness, with statements made by co-partisans being viewed as more truthful.

KEYWORDS

COVID-19, Dirichlet process mixture model, pairwise comparison model, public opinion

1 | INTRODUCTION

The ability to measure latent attributes, as well as differences in how humans perceive these latent attributes, is important for many research areas in the social and behavioural sciences. Examples include research on: perceptions of skin colour (Massey & Martin, 2003), perceptions of racial stereotypicality (Eberhardt et al., 2006), cross-cultural differences in values (Oishi et al., 2005), perceptions of the compactness of legislative districts (Kaufman et al., 2021), understandings of political freedom and political efficacy across cultures (King et al., 2004), and evaluations of biomedical images (Phelps et al., 2015), among many others.

Likert-type scales are one commonly used approach to measure such latent attributes and concepts. However, it is widely understood that human raters may use Likert-type scales differently and this can make it difficult to interpret the resulting estimates (see, for instance, Bachman & O'Malley, 1984; Brady, 1985; Hannon & DeFina 2014; King et al., 2004; Suchman & Jordan, 1990). Furthermore, some approaches using Likert-type scales require the human raters to memorize the relationship between the numbered scale categories and the underlying construct being measured. These cognitive demands can make it more difficult for human raters to use the scale as intended, leading to very low inter-rater reliability (Hannon & DeFina, 2016).

Pairwise comparison methods (Bradley & Terry, 1952; David, 1963; Thurstone, 1927), which elicit binary judgments regarding pairs of items, are less susceptible to these problems. Because these approaches only ask human raters to make a series of binary judgments about which of two paired items has more of the latent attribute of interest, the cognitive demands on the raters are relatively low. Furthermore, there is evidence that approaches using paired comparison data are less likely to suffer from design artefacts due to survey question wording or the labelling of the Likert-type scale categories (Oishi et al., 2005) and are more accurate when compared to objective, gold standard measures (Phelps et al., 2015).

Nonetheless, existing models for paired comparison data are not without their own limitations. Standard models (Bradley & Terry, 1952; Thurstone, 1927) assume that the latent attribute space is unidimensional and that there are no differences in how raters perceive the latent attributes of the paired items. This eliminates the possibility of estimating whether some types of raters perceive the items in question differently than other raters. Understanding such perceptual differences is an important part of recent research in a number of fields (Abrajano et al., 2021; Hannon & DeFina, 2014; Neiss et al., 2009). As discussed in Section 2.1 below, Carlson and Montgomery (2017) include a rater-specific parameter in their model that allows for differences in perceptual sensitivity across raters. While this approach is an improvement, it still assumes that each rater projects the latent item-specific attributes onto the same unidimensional space as all other raters. This can be a serious limitation in applications where the perceptual differences across raters are more extreme. Attempts have been made to develop pairwise comparison models with multidimensional latent attribute spaces (Balakrishnan & Chopra, 2012; Carroll & De Soete,

1991; Yu & Chan, 2001). However, the rater-specific parameters in these approaches are effectively treated as nuisance parameters which again limits what can be learned about rater-specific perceptual differences.

In this article, we address these limitations by developing a new model for pairwise comparisons. The proposed model assumes that each rater's perceptions are a weighted average of the multidimensional latent attributes of the items. This parameterization allows for a straightforward interpretation of the perceptual differences across raters and how those perceptual differences manifest as differences in binary judgments. This is a major advantage over previous multidimensional pairwise comparisons models where interpretation of the rater-specific parameters is not so straightforward.

Moreover, in the second version of the new model, we add a Dirichlet process prior on the rater-specific parameters. This allows us to flexibly cluster raters into groups that share similar perceptual orientations. This ability to group raters by the similarity of their shared perceptions can advance our understanding of how perceptions vary across raters. In addition, examining the associations between rater characteristics and the rater-specific perceptual parameters provides us with information on how perceptual differences vary across identifiable groups of individuals.

We conduct multiple simulation studies with varying amounts of data to demonstrate how the proposed models and model-fitting algorithms work in practice. These studies demonstrate that both versions of the new model are able to accurately estimate the parameter values based on simulated binary choice data. Moreover, we observe that estimation accuracy improves as the datasets become larger.

We apply the new model to original survey data on the perceived truthfulness of public statements made about COVID-19. Our analysis sheds important light on how individuals perceive information on COVID-19 and how their perceptual orientation covaries with their personal characteristics and political beliefs.

We hypothesize that perceptions of truthfulness are influenced by two distinct attributes of the statements: (1) the objective truthfulness of the statements, and (2) the political valence of the statements. Applying our new model to our survey data, we find that the objective truthfulness of a statement only weakly correlates with survey respondent perceptions of truthfulness. On the other hand, we find strong correlations between the political valence of the statements and their perceived truthfulness, moderated by the political leaning of the respondent. Statements made by a co-partisan of a respondent tend to be viewed as more truthful by that respondent. A sizable fraction of respondents gauge the truthfulness of COVID-19 statements through partisan lenses. For these respondents, partisanship has a stronger impact on their responses than does the actual truthfulness of the statements. Indeed, the responses from the most right-leaning respondents are negatively correlated with the objective truthfulness of the statements. That said, a plurality of respondents are relatively unswayed by partisanship but have a difficult time accurately gauging the truthfulness of COVID-19 statements. We also observe associations between the respondent-specific perceptual parameters and public-health-relevant behaviours, such as mask wearing and social distancing.

The remainder of this paper proceeds as follow. First, we detail the new multidimensional pairwise comparison model. Next, we summarize the results from simulation studies of the model. Third, we describe the COVID-19 survey design and data collection procedure. Next, we apply the new model to the pairwise comparison data collected in the survey. We report and compare the results from existing unidimensional models and the newly proposed multidimensional model. We conclude by discussing our findings.

2 | A NEW MODEL FOR PAIRWISE COMPARISONS DATA WITH HETEROGENEOUS PERCEPTIONS

Traditional models for pairwise comparisons assume that objects have unidimensional latent attributes (Bradley & Terry, 1952; David, 1963; Thurstone, 1927). Some researchers add a unidimensional respondent-specific parameter to account for respondents' different levels of ability or sensitivity (Carlson & Montgomery, 2017). There have also been attempts to generalize pairwise comparison models to multidimensional latent spaces (Balakrishnan & Chopra, 2012; Carroll & De Soete, 1991; Yu & Chan, 2001). In this section, we briefly review the existing pairwise comparison models, before we introduce our new model.

2.1 | Existing models

We start by reviewing unidimensional pairwise comparison models. Consider a set of J objects $\{o_j\}_{j=1}^{J}$. We assume that each o_j has a latent attribute $\theta_j \in \mathbb{R}$ that denotes an attribute of interest. While $\boldsymbol{\theta}$ is unobserved, we do observe $y_{ijj'}$, the result of a paired comparison of o_j and $o_{j'}$ by respondent i, in which i is asked to make a ranking judgment as to whether o_j or $o_{j'}$ has a larger value of the latent attribute. $y_{ijj'}$ is equal to 1, if respondent i judges o_j to have a larger value of the latent attribute in question than $o_{j'}$, 0 otherwise. More specifically, we assume

$$y_{ijj'} \sim \mathit{Bernoulli}(p_{ijj'})$$
$$p_{ijj'} = F(\theta_j - \theta_{j'})$$

where $F(\cdot)$ is a cumulative distribution function. If $F(\cdot)$ is the standard normal distribution, the model above is the Thurstone model (Thurstone, 1927). If $F(\cdot)$ is the logistic distribution, the model above is the Bradley–Terry model (Bradley & Terry, 1952). A variant of the above model is to assume that respondents vary in their ability or sensitivity to discern the latent differences between objects, but the latent object attributes remain on the real line, that is, $\theta_j \in \mathbb{R}$ for $j = 1, \ldots, J$. Here

$$p_{ijj'} = F\left(\beta_i \left[\theta_j - \theta_{j'}\right]\right)$$

Typically, it is assumed that $\beta_i \in \mathbb{R}_+$ for $i = 1, \ldots, N$ (Carlson & Montgomery, 2017).

In addition to the unidimensional models, researchers have also proposed multidimensional pairwise comparison models (Cattelan, 2012). For the multidimensional models, both objects and respondents are assumed to have locations in a common latent space. Between any two objects, a respondent prefers the object whose latent location is closer to that of her own. The d-dimensional wandering vector model is a typical multidimensional pairwise comparison model (Carroll & De Soete, 1991), the setup of which is as follows:

$$p_{ijj'} = F(\boldsymbol{\beta}_i \cdot \boldsymbol{\theta}_j - \boldsymbol{\beta}_i \cdot \boldsymbol{\theta}_{j'})$$

where $\boldsymbol{\beta}_i \in \mathbb{R}_+^d$ and $\|\boldsymbol{\beta}_i\| = 1$ for $i = 1, \ldots, N$, and $\boldsymbol{\theta}_j \in \mathbb{R}^d$. A respondent vector, $\boldsymbol{\beta}_i$, is a unit-length vector with non-negative elements. No estimation method was provided for this model when it was originally proposed. Later, MCMC methods have been proposed for the wandering vector model (Balakrishnan & Chopra, 2012; Yu & Chan, 2001). However, these approaches do not

constrain a respondent vector as a unit-length non-negative vector. Instead, they assume a multivariate normal prior for all $\boldsymbol{\beta}_i$, such as $\boldsymbol{\beta}_i \sim N_d(\mathbf{1}, \boldsymbol{I}_d)$. These weaker constraints on respondent attributes lessen the interpretability of the model, since the dot product of a respondent vector and an object vector is no longer the weighted average of an object's attributes on different dimensions.

2.2 | A new multi-dimensional model

The unidimensional pairwise comparison models discussed above have important limitations. They either assume no perceptual differences between respondents, or they assume that respondents only vary in the ability or sensitivity to discern the object attribute differences. Moreover, the unidimensional attribute assumption is overly strong when respondents evaluate objects on more than one latent dimension. Furthermore, respondents may differentially weight the attributes that correspond to different latent dimensions.

Existing multidimensional pairwise comparison models are difficult to interpret due to their lack of constraints on the respondent-specific parameters. When respondent-specific parameters are not constrained to be unit-length non-negative vectors, these parameters cannot be easily viewed as dimension-specific weights. In addition, the existing models do not allow for any clustering among the respondent-specific parameters that would represent shared perceptual frameworks among respondents. To address these issues, we propose a new multidimensional pairwise comparison model. We detail two versions of this model—each corresponding to a different prior distribution for the respondent-specific parameters.

In this new model, we operationalize a unit-length weight vector for each respondent with trigonometric functions. This allows us to model a respondent's perception of an object as the weighted average of the object's attributes on each latent dimension. The model therefore allows researchers to estimate how multiple latent sub-attributes are aggregated into a general latent attribute, and to assess the extent to which respondents differ in their construction of the general attribute from the sub-attributes.

In the first version of the model we assume a uniform prior for these respondent-specific parameters. In the second version, we assume a Dirichlet process prior on the respondent-specific parameters. This second model allows researchers to learn how perceptual frameworks cluster among respondents and how various respondent characteristics relate to respondent perceptions of the latent attributes in interest.

We begin with the special case of a two-dimensional latent attribute space. Once again, consider a set of J objects $\{o_j\}_{j=1}^{J}$. However, we now assume each o_j has latent attributes that can be represented by a location in two-dimensional Euclidean space: $\boldsymbol{\theta}_j \in \mathbb{R}^2$. We assume that respondents differ in the weights they place on each of these two dimensions. More specifically, respondent i's judgment depends on a unit-vector $\mathbf{g}(\gamma_i) \equiv (\cos(\gamma_i), \sin(\gamma_i))^T$ with $\gamma_i \in [0, \frac{1}{2}\pi]$ in the following way:

$$y_{ijj'} \sim \mathit{Bernoulli}(p_{ijj'})$$
$$p_{ijj'} = \Phi_1(\boldsymbol{\theta}_j \cdot \mathbf{g}(\gamma_i) - \boldsymbol{\theta}_{j'} \cdot \mathbf{g}(\gamma_i))$$

where $\Phi_1(\cdot)$ is the CDF of a univariate standard normal distribution, and $\cdot$ denotes the dot product between two vectors. Intuitively, respondent i projects the latent attributes onto $\mathbf{g}(\gamma_i)$ and then uses the signed distance between the projected points to compare two objects. This is depicted graphically in Figure 1.

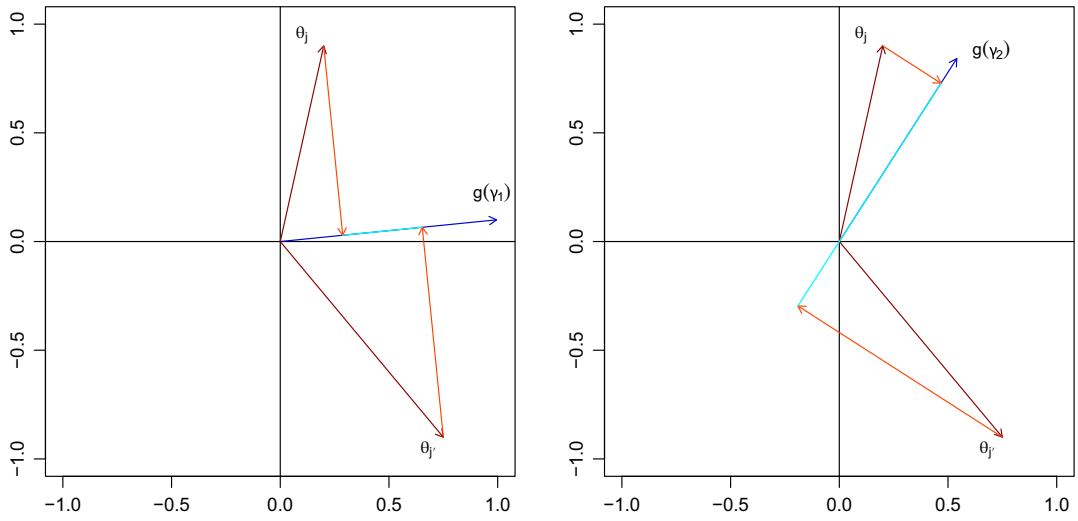

FIGURE 1 Example of the two-dimensional latent attribute model. In the left panel, respondent 1 places much more emphasis on dimension 1 (the horizontal dimension). As a result, this individual is slightly more likely to evaluate $o_{j'}$ as being preferred to o_j. In the right panel, respondent 2 gives weight to both of the latent dimensions with slightly more weight placed on dimension 2 (the vertical dimension). As a result, this individual is much more likely to evaluate o_j as being preferred to $o_{j'}$ [Colour figure can be viewed at wileyonlinelibrary.com]

A semi-conjugate prior distribution for $\boldsymbol{\theta}_j$ is:

$$\boldsymbol{\theta}_j \overset{iid}{\sim} \mathcal{N}_2(\mathbf{0}, \mathbf{I}_2) \quad j = 1, \ldots, J.$$

A number of priors for $\boldsymbol{\gamma}$ are reasonable. We consider two, and introduce the two versions of the new model in the following subsections.

2.2.1 | Uniform prior

The most intuitive prior on γ_i is the uniform prior:

$$\gamma_i \overset{iid}{\sim} \mathcal{U}nif\left(0, \frac{1}{2}\pi\right) \quad i = 1, \ldots, N.$$

This specification has the advantage of simplicity but it does not allow for the possibility that γ_i is equal to $\gamma_{i'}$ for any two individuals i and i'. Such grouping may be desirable if we are interested in making inferences about the extent to which respondents share the same perceptual framework for evaluating the latent attributes in question. Furthermore, allowing γ_i to equal $\gamma_{i'}$ with positive probability is also useful in situations where respondents only rate a small-to-moderate number of paired comparisons. In these situations, allowing some form of clustering among the γ parameters will lower the variance of the resulting estimates of the γ parameters.

2.2.2 | Dirichlet process prior

An alternative is to assume that each γ_i is drawn from a distribution G that is itself drawn from a Dirichlet process. More formally,

$$\gamma_i \overset{iid}{\sim} G \quad i = 1, \ldots, N$$
$$G \sim \mathcal{DP}(\alpha G_0)$$

where $\alpha \in \mathbb{R}_+$ is a concentration parameter and G_0 is the centring distribution, which is specified as $\mathcal{U}nif(0, \frac{1}{2}\pi)$. α could be either fixed at a constant value or given a prior distribution and estimated. If α is to be estimated, then we assume a Gamma prior distribution for α:

$$\alpha \sim \mathcal{G}amma(a, b)$$

While the Dirichlet process prior for γ complicates estimation, it has the advantage of allowing for the possibility of perceptual clustering among respondents.

The new models can be generalized to $d > 2$ dimensions by assuming that each $\boldsymbol{\theta}_j \in \mathbb{R}^d$, and the perceptual unit-vectors $\mathbf{g}(\boldsymbol{\gamma}_i)$ are constrained to lie in the positive orthant of $\mathbb{R}^d$. $\boldsymbol{\gamma}_i$ would be $(d-1)$-dimensional in this case with each element being an angle in the positive orthant. For example, if $d = 3$, then we can use $[\gamma_{i1}\ \gamma_{i2}]^T$ to represent a unit-length vector in the positive orthant, $[\cos(\gamma_{i1})\cos(\gamma_{i2}),\ \cos(\gamma_{i1})\sin(\gamma_{i2}),\ \sin(\gamma_{i1})]^T$. The MCMC algorithms we use for fitting these two versions of the model are discussed in Appendix A.

3 | SIMULATION STUDY

We conduct simulation studies to illustrate how the samplers for the two versions of the new model work. The experiments show that both versions of the new model are able to recover the true latent variable values from the observed binary choice data. For both versions of the new model, we specify four configurations of respondent number, I, and object number, J: $(I = 40, J = 40)$, $(I = 40, J = 80)$, $(I = 80, J = 40)$, $(I = 80, J = 80)$. For each configuration, we repeat the simulation steps 50 times, resulting in 50 simulated data sets for each configuration. We use two measures to gauge how well the model can uncover the true latent variables values: the correlations between the estimated parameters and the true values, and the mean squared error (MSE) of the estimated parameters. We compute the correlations and MSEs for the results from each simulation data set under each simulation configuration.

The results of the simulation studies of the first model with the uniform prior on $\boldsymbol{\gamma}$ can be summarized as follows. In the simulations with the least information ($I = 40$ and $J = 40$), the modal correlation between the estimated γ's and their true values is approximately 0.85 and the mode of the MSEs is approximately 0.09. As J increases from 40 to 80, the modal correlation between the estimated γ's and their true values increases to approximately 0.95 and the modal MSE value for these parameters drops to about 0.03. Increasing the number of objects to be rated (J) does more to increase the precision of the estimated $\boldsymbol{\gamma}$ than increasing the number of raters (I). In the simulations with $I = 40$ and $J = 40$, the modal correlations between the estimated θ's and their true values are around 0.9, and the mode of the MSEs is around 0.25. As I increases to 80 and J increases to 80, the modal correlation between the estimated θ's and their true values increases to approximately 0.97 and the modal MSE value for these parameters drops to about 0.07. The simulations suggest that accurate estimation of $\boldsymbol{\theta}$ depends, to a greater extent, on both the number of raters (I) and the number of objects being rated (J) than does accurate estimation of $\boldsymbol{\gamma}$ which is more heavily dependent on J.

The summary for the simulation study on the second version of the new model is as follows. Under the simulation configuration with $I = 40$ and $J = 40$, the modal correlation between the estimated γ values and their true values is around 0.88 and the mode of the MSEs is approximately 0.07. As J increases from 40 to 80, the modal correlation between the estimated γ's and their true values increases to approximately 0.99 and the modal MSE value for these parameters drops to about 0.01. Once again, we observe that increasing the number of objects to be rated (J) has a greater impact on the precision of the $\boldsymbol{\gamma}$ estimates than increasing the number of raters (I). In the simulations with $I = 40$ and $J = 40$, the modal correlations between the estimated θ's and their true values are around 0.9, and the mode of the MSEs is around 0.25. As I increases to 80 and J increases to 80, the modal correlation between the estimated θ's and their true values increases to approximately 0.97 and the modal MSE value for these parameters drops to about 0.05. We again observe that accurate estimation of $\boldsymbol{\theta}$ depends on both I and J to a greater extent than does accurate estimation of $\boldsymbol{\gamma}$.

In sum, across all the simulation repetitions, we consistently observe high correlations between estimated parameters and the true values, and the MSEs of the estimated parameters are low relative to the standard deviation. Moreover, as I and J increase, the samplers for both versions of the new model perform better at latent variable estimation. The detailed simulation study results are reported in the supplemental information document.

4 | APPLICATION: THE PERCEIVED TRUTHFULNESS OF PUBLIC COVID-19 STATEMENTS

In this section we apply our new model to the substantive application of how survey respondents perceive the truthfulness of public statements about COVID-19. Replication data and code are archived at the Harvard Dataverse, https://doi.org/10.7910/DVN/KBAJJO.

4.1 | COVID-19 statements

Since we are interested in the extent to which members of the mass public accurately assess the truthfulness of statements about COVID-19, it is important that we use fact-checked statements so as to have an independent measure of the truthfulness of each statement. Our source of these fact-checked COVID-19 statements is the website https://www.politifact.com. PolitiFact's Editor-in-Chief, Angie Drobnic Holan, gave us permission to use the PolitiFact data for this survey in an email on 11 May 2020.

PolitiFact catalogues a range of statements that have political content. According to PolitiFact's own website: 'Each day, PolitiFact journalists look for statements to fact check. We read transcripts, speeches, news stories, press releases and campaign brochures. We watch TV and scan social media. Readers send us suggestions via email to `truthometer@politifact.com`; we often fact-check statements submitted by readers. Because we cannot feasibly check all claims, we select the most newsworthy and significant ones (Holan, 2020)'. PolitiFact journalists fact check these statements and categorize the truthfulness of each statement into one of six categories (from most truthful to least truthful): true, mostly true, half true, mostly false, false, pants on fire (Holan, 2020).

We selected 42 statements with the intent of balancing the truthfulness of the statements and the slant of the statements (left, neutral and right). These statements were made between 22 February 2020 and 8 May 2020. Ideally, we would have used equal numbers of left-, neutral-

and right-leaning statements from all six truthfulness categories. However, some categories were sparsely populated and we were forced to dichotomize the truthfulness categories into high truth (true, mostly true and half true) and low truth (pants on fire, false and mostly false). This gave us seven statements in each of the 3 × 2 combinations of slant × truthfulness. The full set of 42 statements along with their truthfulness ratings and slant is presented in the supplemental information document.

4.2 | The survey

The survey was conducted on 8 July 2020. Respondents were recruited from the Lucid Marketplace (https://luc.id/marketplace/). Quotas were used to make the sample approximate the US voting age population. The survey was conducted online using the Qualtrics interface.

In this survey, respondents were asked to report their view of the relative truthfulness of COVID-19 statements given to them in randomly selected pairs of statements. Figure 2 depicts what this looks like for one randomly selected pair of statements. After the paired comparisons of COVID-19 statements were given to respondents, the respondents were asked a sequence of demographic, attitudinal and behavioural questions.

We removed 187 respondents that Lucid flagged as having a high likelihood of being fraudulent. We received usable responses from 2,621 respondents. On average, each respondent gave us their view of the relative truthfulness of just less than 15 pairs of randomly selected statements. We provide more information about our survey sample in Appendix B. In addition, the Supplemental Information provides descriptive statistics on our respondent sample.

4.3 | Results

Before presenting results from our new model, we present results from simple unidimensional models. By comparing and contrasting the results from the simple unidimensional models

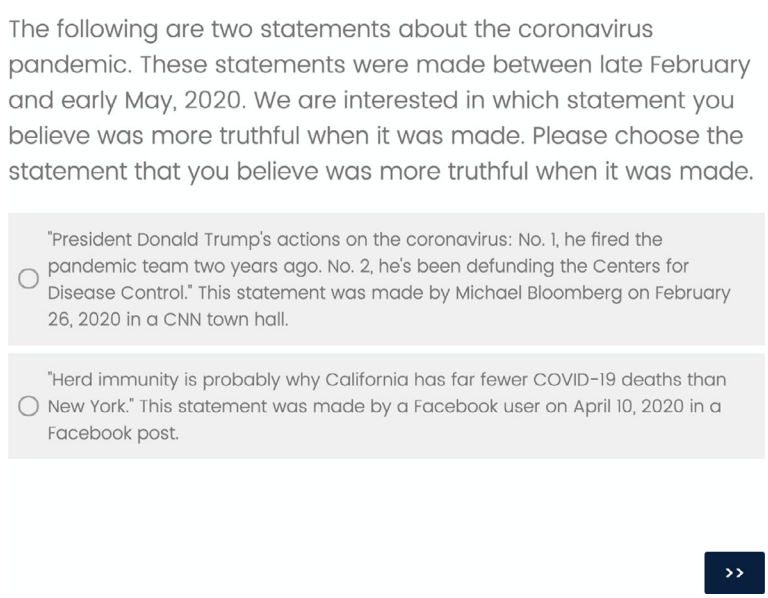

FIGURE 2 Screen shot of COVID-19 statement comparison survey [Colour figure can be viewed at wileyonlinelibrary.com]

and the new model, we show that the new model leads to more interpretable and insightful findings.

4.3.1 | Results from unidimensional models

As a starting point, we fit the simple Thurstone model

$$y_{ijj'} \sim \mathcal{B}ernoulli(p_{ijj'})$$
$$p_{ijj'} = \Phi(\theta_j - \theta_{j'})$$

to the pairwise comparisons data from our survey. Here $\Phi(\cdot)$ is the CDF of a standard normal distribution. To identify the model, we constrained θ_{1014} to be negative (statement 1014 is a neutral-valence, low-truth statement) and constrained $\theta_{1015} = 0.25$ (statement 1015 is a neutral-valence, high-truth statement). These constraints are consistent with an interpretation of the latent dimension as objective truthfulness. The remaining θ parameters were assumed to have independent standard normal prior distributions. The MCMC sampler was run for 120,000 iterations with the first 20,000 discarded as burn-in iterations. Every 10th iteration was stored.

Inspection of the output reveals that this simple model provides a poor fit to the observed data. For instance, for each observed $y_{ijj'}$ we calculate the in-sample posterior expectation of a correct classification:

$$\frac{1}{M}\sum_{m=1}^{M}\left(\mathbb{I}(y_{ijj'}=1)\Phi\left(\theta_j^{(m)}-\theta_{j'}^{(m)}\right)+\mathbb{I}(y_{ijj'}=0)\Phi\left(\theta_{j'}^{(m)}-\theta_j^{(m)}\right)\right)$$

where $m = 1, \ldots, M$ indexes the MCMC draws. Note that a 'correct' classification is simply defined to be a classification equal to the observed response—it is not necessarily related to whether respondent i accurately perceived the true truthfulness of statement j relative to statement j'.

The average of these posterior expectations of a correct response, taken over all observed $y_{ijj'}$s, is 0.52. We can also aggregate to the statement by averaging over respondents. Doing this, we see that the average probability of a correct classification across all statements is also 0.52, and that no statement has a probability of being correctly classified greater than 0.56. If we aggregate to the respondent by averaging over the statement pairs seen by each respondent, we see that the average probability of a correctly classified response by a respondent is also 0.52. Furthermore, we find that 26% of respondents have probabilities of a correctly classified statement less than 0.5 and only 0.3% of respondents (8 of 2,621) have probabilities of a correctly classified statement greater than 0.6.

We also examine how the posterior means of the θ parameters correlate with the objective truth and partisan valence of the statements. To do this we give a 'pants-on-fire' statement a value of 0, a 'false' statement a value of 1, a 'mostly false' statement a value of 2, a 'half-true' statement a value of 3, a 'mostly true' statement a value of 4 and a 'true' statement a value of 5. We then calculated the Spearman rank correlation between these truthfulness ratings and the posterior means of the θ parameters. This produced a rank correlation of 0.42.

Similarly, we gave right-valence statements a value of 1, neutral-valence statements a value of 0, and left-valence statements a value of −1. Then we calculated the Spearman rank correlation between the partisan valence of the statements and the posterior means of the θ parameters. This resulted in a rank correlation of −0.17.

The simple unidimensional Thurstone model produces estimates of the statement-specific parameters that are only weakly correlated with objective truth and even more weakly correlated with the other factor that we expect to structure responses—the political valence of the statements.

As noted above, a natural extension of the basic Thurstone model is to introduce a respondent-specific parameter β_i that allows for differential ability to perceive differences between statements. This produces the model:

$$y_{ijj'} \sim \mathcal{B}ernoulli(p_{ijj'})$$
$$p_{ijj'} = \Phi(\beta_i[\theta_j - \theta_{j'}]).$$

We fit this model to the pairwise comparisons data from our survey. To identify the model, we again constrained θ_{1014} to be negative and constrained $\theta_{1015} = 0.25$. Again, these constraints are consistent with the latent dimension being interpreted as objective truthfulness. The remaining θ parameters were assumed to have independent standard normal prior distributions. The β parameters were also assumed to have independent standard normal priors. The sign of the β parameters was not restricted. The MCMC sampler was run for 120,000 iterations with the first 20,000 discarded as burn-in iterations. Every 10th iteration was stored.

If we calculate the in-sample posterior expectation of a correct classification for this model in the analogous way that we did for the simple Thurstone model, we find that the average of these posterior expectations of a correct response, taken over all observed $y_{ijj'}$s, is 0.55. While the inclusion of the respondent-specific β parameters ensures that the respondent-level predictions match the observed data at least 50% of the time, it is still the case that 35% of the observed $y_{ijj'}$s have posterior probabilities of a correct classification less than 0.50. At the statement level, we see that, on average, statements are classified correctly 55% of the time with only 2 of 42 statements having a probability of correct classification greater than 0.6.

The Spearman rank correlation between the posterior means of the statement-specific θ parameters and the objective truthfulness of the statements is 0.32, which is lower than in the simple Thurstone model. However, the rank order correlation between the posterior means of θ and the partisan valence of the statements is -0.73. Even though we constrained the model so that a neutral-valence, high-truth statement was to the right of 0 and a neutral-valence, low-truth statement was to the left of 0, the resulting estimates of θ are more strongly correlated with the partisan valence of the statements than the objective truthfulness of the statements.

Indeed, this warping of truthfulness at the respondent level can be seen in the posterior means of the respondent-specific β parameters. 38% of respondents have a β parameter with a posterior mean less than 0. In other words, 38% of respondents are, on average, viewing objective truth as subjective falsity, and vice versa.

4.3.2 | Results from the two-dimensional Dirichlet process model

The results from the simple unidimensional models are not fully satisfying. The Thurstone model does a poor job of representing observed patterns in the data, and produces estimates of statement-specific parameters that only weakly correlate with objective truthfulness. The inclusion of a respondent-specific parameter slightly improves model fit at the expense of weakening

the already weak correlation between the statement-specific parameter estimates and objective truth.

We fit the two-dimensional Dirichlet process model discussed in Section 2.2.2 to the data in the hope that this provides a better fit than the unidimensional models. To identify the model, we constrained θ_{1015} to be equal to 0.25 on the first dimension and greater than 0 on the second dimension (statement 1015 is a neutral-valence, high-truth statement); we constrained θ_{1004} to be less than 0 on the first dimension (statement 1004 is a right-valence, low-truth statement); and we constrained θ_{1042} to be greater than 0 on the first dimension (statement 1042 is a left-valence, high-truth statement). The remaining θ parameters were assumed to have independent bivariate normal prior distributions with mean $\mathbf{0}$ and variance–covariance matrices equal to identity matrices. G_0 was set to $\mathcal{U}nif(0, \frac{1}{2}\pi)$ and the concentration parameter α was assumed to follow a $\mathcal{G}amma(1, 1)$ distribution. The MCMC sampler was run for 440,000 iterations with the first 40,000 discarded as burn-in iterations. Every 40th iteration was stored.

Calculating the in-sample posterior expectation of a correct classification for this model in the analogous way that we did for the unidimensional models above, we find that the average of these posterior expectations of a correct response, taken over all observed $y_{ijj'}$s, is 0.57. The respondent-level predictions and statement-level predictions also match the observed data 57% of the time. These numbers are slightly better than the values of 0.52 and 0.55 we achieved with the unidimensional models.

However, this slight improvement in in-sample predictive accuracy is not the main advantage of the two-dimensional Dirichlet process model. The main advantage is that it allows us to uncover a more nuanced understanding of how certain types of respondents assess the truthfulness of statements. More specifically, in this application it allows us to see how some respondents make assessments of truthfulness based on their partisanship or political ideology, while other respondents seem to be more guided by the objective truthfulness of the statements.

As a starting point, consider Figure 3. This figure plots the posterior means of θ_j for the $j = 1, \ldots, 42$ statements along with $\mathbf{g}(0.18)$, $\mathbf{g}(0.77)$, and $\mathbf{g}(1.41)$, where 0.18, 0.77, and 1.41 are the minimum, median, and maximum values of the posterior means of γ_i for $i = 1, \ldots, N$. Figure 3 allows us to see how three types of respondents (those with $\gamma_i = 0.18, 0.77$ and 1.41) perceive the truthfulness of the statements.

Respondents with γ_i parameters near the median of 0.77 project the statement-specific θ parameters onto a dimension that correlates with the objective truthfulness of the statements, albeit weakly. On the other hand, respondents with γ_i parameters near the minimum of 0.18 project the statement-specific θ parameters onto a dimension which is positively correlated with the leftward valence of the statements. Finally, respondents with γ_i parameters near the maximum of 1.41 project the statement-specific θ parameters onto a dimension which is positively correlated with the rightward valence of the statements.

While the information in Figure 3 is useful, it does not provide information on three important things: (a) the distribution of respondent-specific γ_i parameters, (b) the precise strength of the correlation between particular $\mathbf{g}(\gamma_i)$ projections and the objective truthfulness of statements and (c) the precise strength of the correlation between particular $\mathbf{g}(\gamma_i)$ projections and the left–right valence of the statements. This information is displayed in Figure 4.

Looking at Figure 4, three things are apparent. First, most respondents have estimated γ parameters near the median posterior mean value of 0.77 (the modal estimate of γ_i is just to the right of 0.77). There are a smaller number of respondents who have lower estimated γ parameters—about 16% of the respondents have estimated γ parameters less than 0.5—and there

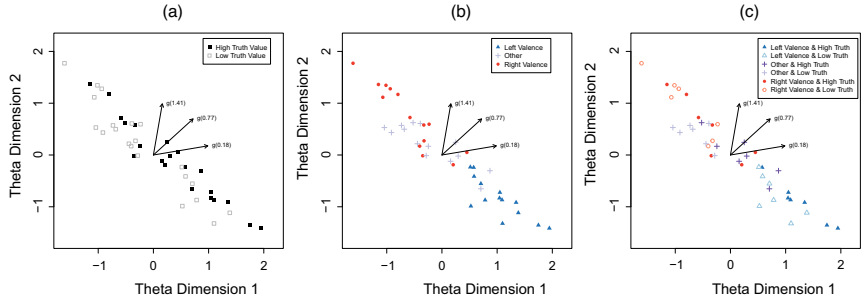

FIGURE 3 Posterior Means of $\boldsymbol{\theta}$ and the minimum, maximum, and median posterior means of $\mathbf{g}(\boldsymbol{\gamma})$. In each panel, the points correspond to the posterior means of the $\boldsymbol{\theta}$ parameters for the 42 statements. The arrows correspond to the $\mathbf{g}(\boldsymbol{\gamma})$ vectors at the minimum posterior mean of γ_i, the maximum posterior mean of γ_i, and the median posterior mean of γ_i for $i = 1, \ldots, N$. In panel (a), the $\boldsymbol{\theta}$ points are shaded based on the objective truthfulness of the statements. In panel (b), the $\boldsymbol{\theta}$ points are colour-coded based on the left–right valence of the statements. Finally, in panel (c), the $\boldsymbol{\theta}$ points are coded according to both the objective truthfulness of the statements and the left–right valence of the statements. Note that projecting the $\boldsymbol{\theta}$ points onto $\mathbf{g}(0.77)$ (0.77 is the median posterior mean of $\gamma_i \quad i, \ldots, N$) produces values associated with the objective truthfulness of the statements, albeit weakly. On the other hand, projecting the $\boldsymbol{\theta}$ points onto $\mathbf{g}(0.18)$ and $\mathbf{g}(1.41)$ results in points where higher values correspond to more left-leaning and right-leaning valence respectively [Colour figure can be viewed at wileyonlinelibrary.com]

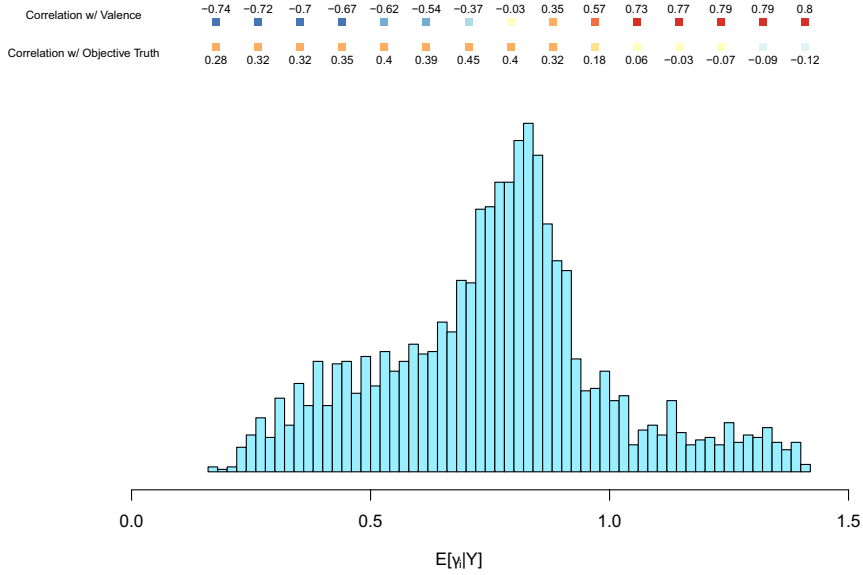

FIGURE 4 Histogram of the posterior means of γ_i for $i = 1, \ldots, N$ along with the spearman rank correlations between $\boldsymbol{\theta}_j \cdot \mathbf{g}(\gamma)$ and objective truthfulness and left–right valence for various values of γ. Note that respondents whose posterior mean γ parameter is near 0.77 tend to assess statements primarily based on the objective truthfulness of the statements but that this association is weak (correlation slightly greater than 0.4). Respondents with γ parameters that are closer to the extremes of 0.18 and 1.41 assess the truthfulness of the COVID-19 statements in ways that are strongly associated with the left–right valence of the statements. Further, respondents with γ parameters greater than approximately 1.1 not only assess the truthfulness of the COVID-19 statements such that right-valence statements are perceived as more truthful, they also assess the truthfulness of COVID-19 statements in ways that are negatively correlated with the objective truthfulness of the statements. All correlations were calculated across all $J = 42$ statements [Colour figure can be viewed at wileyonlinelibrary.com]

are an even smaller number of respondents with much larger estimated γ parameters—about 13% have estimated γ parameters greater than 1.0.

Second, Figure 4 presents the correlation between $\boldsymbol{\theta}_j \cdot \mathbf{g}(\gamma)$ and the objective truthfulness of the statements for various values of γ, for $j = 1, \ldots, 42$. This is the multidimensional analogue to the correlation between $\boldsymbol{\theta}$ and objective truthfulness from the unidimensional models discussed in Section 4.3.1. Once again, we give a 'pants-on-fire' statement a value of 0, a 'false' statement a value of 1, a 'mostly false' statement a value of 2, a 'half-true' statement a value of 3, a 'mostly true' statement a value of 4 and a 'true' statement a value of 5. We then calculated the Spearman rank correlation between these truthfulness ratings and $\boldsymbol{\theta}_j \cdot \mathbf{g}(\gamma)$ for the posterior means of $\boldsymbol{\theta}$ for the 42 statements and 15 equally spaced values of gamma from 0.18 to 1.41. This produces the 15 colour-coded correlations at the top of Figure 4.

What we see here is that the γ value that induces the highest correlation with the objective truth of the statements is $\gamma = 0.70$ which gives rise to a correlation of 0.45. γ values less than or equal to 0.88 give rise to correlations with objective truth that are greater than or equal to 0.28. However, individuals with γ values greater than or equal to 1.14 tend to rate COVID-19 statements in ways that are negatively correlated with the objective truth of the statements.

Third, Figure 4 presents the correlation between $\boldsymbol{\theta}_j \cdot \mathbf{g}(\gamma)$ and the left–right valence of the statements. As above, we gave right-valence statements a value of 1, neutral-valence statements a value of 0, and left-valence statements a value of −1. We calculated the Spearman rank correlation between these left–right valence ratings and $\boldsymbol{\theta}_j \cdot \mathbf{g}(\gamma)$ for the posterior means of $\boldsymbol{\theta}$ for the 42 statements and 15 equally spaced values of gamma from 0.18 to 1.41. This produces the 15 colour-coded correlations at the very top of Figure 4.

The resulting pattern of correlations with the left–right valence is stark. Respondents with the highest values of γ, say above or equal to 1.14, perceive the truthfulness of the COVID-19 statements in a way that positively correlates with the rightward valence of the statements with correlations of 0.77 or above. These same individuals' evaluations of the truthfulness of the COVID-19 statements are negatively correlated with objective truth. On the other hand, individuals with the lowest values of γ, say below or equal to 0.35, perceive the truthfulness of the COVID-19 statements in a way that negatively correlates with the rightward valence of the statements with correlations of −0.70 or below. These individuals' evaluations of the truthfulness of the COVID-19 statements are weakly positively correlated with the objective truth of the statements.

To summarize, respondents with the modal value of γ are primarily responding to the objective truthfulness of the COVID-19 statements when evaluating the truthfulness of pairs of statements. These respondents are not rating statements in ways that are correlated with the left–right valence of the statements. That said, there is only a weak correlation between the objective truth of the statements and their subjective perceptions. On the other hand, respondents with γ values at the two extremes are rating the truthfulness of statements in ways that are strongly associated with left–right valence of the statements—respondents with low values of γ tend to see left-leaning statements as more truthful while respondents with high values of γ tend to see right-leaning statements as more truthful. The objective truth of the statements is less relevant for these respondents than is the left–right valence of the statements. Indeed, those respondents who tend to see right-leaning statements as more truthful tend to perceive the truthfulness of the statements in ways that are slightly negatively correlated with the objective truth of the statements.

We also examine how respondent perceptions of COVID-19 statement truthfulness, as measured by their estimated γ parameters, correlate with respondent characteristics and behaviours. Figure 5 plots the relationship between the respondent-specific γ estimates and three measures related to the political attitudes of respondents: partisanship (as operationalized by an indicator

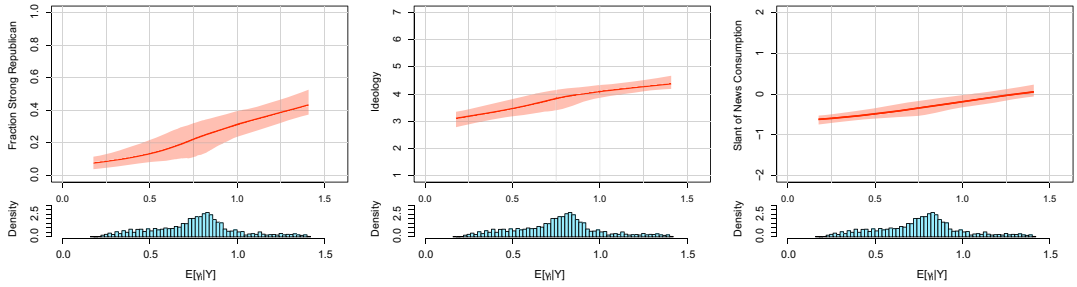

FIGURE 5 Associations between posterior means of γ_i for $i = 1, \ldots, N$ and respondent partisanship, ideology, and slant of news media consumption. The dark orange lines are the posterior means of local regression predictions averaged over the posterior distribution of $\boldsymbol{\gamma}$. The light orange band is the pointwise central 95% credible region for these local regression predictions, again averaged over the posterior distribution of $\boldsymbol{\gamma}$. Each panel of this figure plots a local regression estimate of the conditional expectation function of the variable in question on γ_i for respondents $i = 1, \ldots, N$. Each panel was constructed by fitting M local regressions of the variable in question on each of the M posterior samples of $\boldsymbol{\gamma}$. The pointwise average of these M estimated regression functions is the dark orange line in each panel. The light orange band in each panel is the pointwise central 95% credible region for these local regressions (the empirical 2.5th and 97.5th pointwise percentiles of the M estimated regression functions) [Colour figure can be viewed at wileyonlinelibrary.com]

of whether a respondent self-identifies as a strong Republican), ideology (as operationalized by respondent self-placement on a 7-point Likert-type scale running from 1 = 'very liberal' to 7 = 'very conservative') and the slant of news media consumption (as operationalized by respondent self-statement of their preferred news outlet combined with the media bias ratings from https://www.allsides.com/media-bias/media-bias-ratings).

Not surprisingly, inspection of Figure 5 reveals that right-wing partisanship, ideology and news media consumption is increasing in γ. Respondents with the largest values of γ tend to be the respondents with the most right-leaning political views. Those with the lowest γ values tend to be the most left-leaning respondents.

We also examine whether respondent-specific γ values (and thus the perceptual framework that respondents use to evaluate the truthfulness of COVID-19 statements) are associated with behaviours important for public health. More specifically, Figure 6 plots the relationship between the respondent-specific γ estimates and (a) a measure of a lack of social distancing (operationalized as 0/1 indicator equal to 1 if a respondent said that 21 or more people were 6 feet or closer to them in the past week), and (b) a measure of mask wearing (operationalized as the number of situations, out of nine possible, where the respondents said they wear a mask). The panels are constructed in the same way as Figure 5.

Figure 6 shows that the structure underlying how respondents judge the truthfulness of COVID-19 statements (as measured by their γ values) is associated with behaviours that have consequences for public health. Specifically, lack of social distancing is increasing in γ, while mask wearing is decreasing in γ.

5 | DISCUSSION

In this paper, we have proposed a new pairwise comparison model to measure multidimensional latent attributes and respondent-specific perceptual parameters. This model incorporates interpretable constraints on respondent-specific parameters, and explicitly models how object

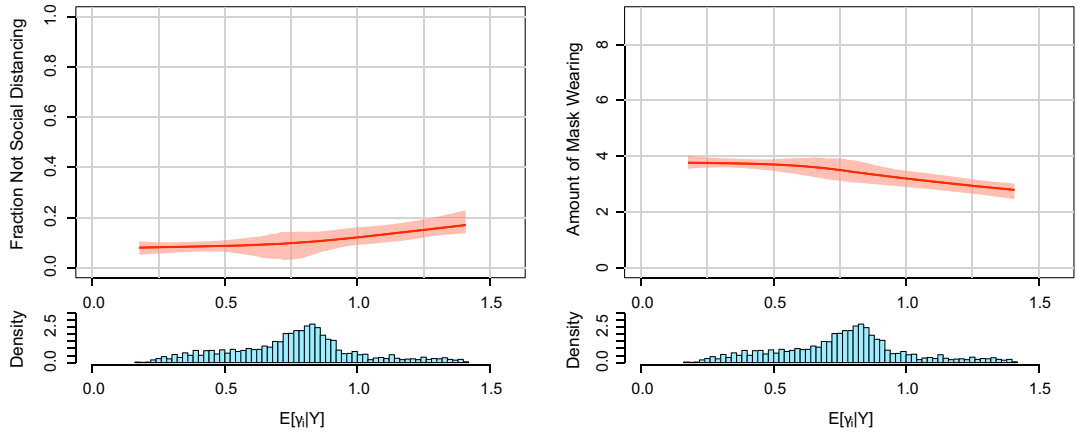

FIGURE 6 Associations between posterior means of γ_i for $i = 1, \ldots, N$ and self-reported social-distancing behaviour and mask-wearing behaviour. The dark orange lines are the posterior means of local regression predictions averaged over the posterior distribution of $\boldsymbol{\gamma}$. The light orange band is the pointwise central 95% credible region for these local regression predictions, again averaged over the posterior distribution of $\boldsymbol{\gamma}$ [Colour figure can be viewed at wileyonlinelibrary.com]

attributes on different dimensions are aggregated into a respondent's choices. The new model improves upon previous models in that it provides an easily interpretable framework for characterizing and estimating respondent-specific differences in perceptions along with multidimensional latent attributes of objects. We fit the model using MCMC methods. Software for fitting the model is freely available in the `MCMCpack` R package (Martin et al., 2011).

To illustrate the strength of the new model, we have applied the new model to both simulated data and original survey data. The simulation studies show that the new model is able to recover the true values of latent variables based on observed binary choice data. In the survey data application, our analysis sheds light on how statements on COVID-19 are perceived by respondents and what respondent characteristics are associated with the perceptual frameworks used by respondents. Importantly, we find a weak correlation between the actual truthfulness of a statement and respondents perceptions of truthfulness. More importantly, we find that the political valence of statements is largely responsible for the variation in perceived truthfulness. Co-partisanship between a respondent and the speaker of a statement predicts higher perceived truthfulness. The respondent-specific parameters estimated in the new model also bear out the general patterns shown in the respondents' perceptions, and the associations between perceptions and behaviours. Our findings show that individuals generally have a hard time differentiating truthful information on COVID-19 from false information. Moreover, many respondents rely on partisanship as a cue to gauge the truthfulness of information on COVID-19. Among these partisan respondents, the most rightward-leaning respondents' tend to view objectively truthful statements as subjectively false. Finally, we also observe associations between the respondent-specific perceptual parameters and respondents' practice of mask wearing or social distancing.

It is important to note that there are limitations to our work. As with many other latent attribute models, proper use of our new models requires subject matter expertise at a number of points. The models are only identifiable after constraints are placed on the model parameters. While the constraints are, to some degree, arbitrary; some choices will result in more easily interpretable results than others. Subject matter expertise should thus inform these decisions. Relatedly, in applications that focus on the rater-specific parameters, the objects selected for

rating determine the estimand and thus affect the results. These decisions should be informed by domain-specific knowledge.

There are also limitations to our COVID application. The conclusions we reach are based on the sample of respondents from July 2020 and the statements they were given to evaluate were from the early days of the pandemic. We are thus only able to make inferences about public perceptions in this time period. We were also limited in the number of statements that we could use. As we saw in our simulation studies, we would have increased the precision of our estimates if we were able to use more statements. Finally, we only looked at public perceptions of the truthfulness of statements about COVID-19. Accordingly, we are not able to say anything about how these public perceptions are similar to or different from perceptions of statements in other policy areas such as economic policy.

More work is required to fully realize the potential of our proposed models. First, additional work is warranted on the question of how to most efficiently allocate pairs of items to the raters. We suspect that an active learning approach may be dramatically more efficient than the simple randomization scheme that we used in our survey. Second, while it is clear how to extend our model to latent spaces with three or more dimensions, efficiently fitting such extended models may require modifications to our MCMC algorithm. Finally, we think there is room for more work on model evaluation within this class of models.

ACKNOWLEDGEMENTS

Quinn's research was supported by the U.S. National Science Foundation (grant SES 16-59922). The survey was funded by the University of Michigan.

DATA AVAILABILITY STATEMENT

We have added the functions implementing the new model to the "MCMCpack" R package. We cited this R package in the conclusion section and in the Appendix. The code is now available on CRAN. We have also hosted the replication data and code in a Harvard Dataverse repository. It can be accessed here: https://urldefense.com/v3/__https://doi.org/10.7910/DVN/KBAJJO__;!!N11eV2iwtfs!vLl4T8uK9UTbafiRG3tuesmjmbYYzfLL2h8NR80WddPeKqXfDTXjlO9TZQBmHTBQQTnm-6NXgMNLPrU7Jw$. We have noted this in the manuscript at the beginning of the application section.

ORCID

Qiushi Yu https://orcid.org/0000-0003-3011-4765

REFERENCES

Abrajano, M.A., Elmendorf, C.S. & Quinn, K.M. (2021) Measuring perceived skin color: Spillover effects and likert-type scales. Working Paper.

Albert, J.H. & Chib, S. (1993) Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(June), 669–679.

Bachman, J.G. & O'Malley, P.M. (1984) Yea-saying, nay-saying, and going to extremes: black-white differences in response styles. *Public Opinion Quarterly*, 48(2), 491–509.

Balakrishnan, S. & Chopra, S. (2012) Two of a kind or the ratings game? Adaptive pairwise preferences and latent factor models. *Frontiers of Computer Science*, 6(2), 197–208.

Bradley, R.A. & Terry, M.E. (1952) Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4), 324–345.

Brady, H.E. (1985) The perils of survey research: inter-personally incomparable responses. *Political Methodology*, 11(June), 269–290.

Carlson, D. & Montgomery, J.M. (2017) A pairwise comparison framework for fast, flexible, and reliable human coding of political texts. *American Political Science Review*, 111(4), 835–843.
Carroll, J.D. & De Soete, G. (1991) Toward a new paradigm for the study of multiattribute choice behavior: spatial and discrete modeling of pairwise preferences. *American Psychologist*, 46(4), 342.
Cattelan, M. (2012) Models for paired comparison data: a review with emphasis on dependent data. *Statistical Science*, 27, 412–433.
Chib, S. & Greenberg, E. (1995) Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(November), 327–336.
Connor, R.J. & Mosimann, J.E. (1969) Concepts of independence for proportions with a generalization of the Dirichlet distribution. *Journal of the American Statistical Association*, 64(325), 194–206.
David, H.A. (1963) *The method of paired comparisons*, Volume Number 12 of Griffin's Statistical Monographs and Courses. New York: Hafner Publishing Company.
Eberhardt, J.L., Davies, P.G., Purdie-Vaughns, V.J. & Johnson, S.L. (2006) Looking deathworthy perceived stereotypicality of black defendants predicts capital-sentencing outcomes. *Psychological Science*, 17(5), 383–386.
Escobar, M.D. (1994) Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association*, 89(425), 268–277.
Escobar, M.D. & West, M. (1995) Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90(430), 577–588.
Ferguson, T.S. (1973) A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2), 209–230.
Hannon, L. & DeFina, R. (2014) Just skin deep? The impact of interviewer race on the assessment of african american respondent skin tone. *Race and Social Problems*, 6(4), 356–364.
Hannon, L. & DeFina, R. (2016) Reliability concerns in measuring respondent skin tone by interviewer observation. *Public Opinion Quarterly*, 80(2), 534–541.
Holan, A.D. (2020) The principles of the truth-o-meter: Politifact's methodology for independent fact-checking. Available from: https://www.politifact.com/article/2018/feb/12/principles-truth-o-meter-politifacts-methodology-i/
Ishwaran, H. & James, L.F. (2001) Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453), 161–173.
Ishwaran, H. & Zarepour, M. (2000) Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika*, 87(2), 371–390.
Kaufman, A., King, G. & Komisarchik, M. (2021) How to measure legislative district compactness if you only know it when you see it. *American Journal of Political Science*, 65(3), 533–550.
King, G., Murray, C.J., Salomon, J.A. & Tandon, A. (2004) Enhancing validity and crosscultural comparability of measurement in survey research. *American Political Science Review*, 98, 191–207.
MacEachern, S.N. (1994) Estimating normal means with a conjugate style Dirichlet process prior. *Communications in Statistics-Simulation and Computation*, 23(3), 727–741.
MacEachern, S.N. & Müller, P. (1998) Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics*, 7(2), 223–238.
Martin, A.D., Quinn, K.M. & Park, J.H. (2011) Mcmcpack: Markov chain Monte Carlo in R. *Journal of Statistical Software*, 42(9), 1–21.
Massey, D.S. & Martin, J.A. (2003) The NIS skin color scale. Office of Population Research, Princeton University.
Müller, P. & Rodriguez, A. (2013) Dirichlet process. In Nonparametric Bayesian Inference, pp. 23–41. IMS and ASA.
Neal, R.M. (2000) Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2), 249–265.
Neiss, M.B., Leigland, L.A., Carlson, N.E. & Janowsky, J.S. (2009) Age differences in perception and awareness of emotion. *Neurobiology of Aging*, 30(8), 1305–1313.
Oishi, S., Hahn, J., Schimmack, U., Radhakrishan, P., Dzokoto, V. & Ahadi, S. (2005) The measurement of values across cultures: a pairwise comparison approach. *Journal of Research in Personality*, 39, 299–305.
Phelps, A.S., Naeger, D.M., Courtier, J.L., Lambert, J.W., Marcovici, P.A., Villanueva-Meyer, J.E. et al. (2015) Pairwise comparison versus Likert scale for biomedical image assessment. *American Journal of Roentgenology*, 204(1), 8–14.
Sethuraman, J. (1994) A constructive definition of Dirichlet priors. *Statistica Sinica*, 4, 639–650.
Suchman, L. & Jordan, B. (1990) Interactional troubles in face to face survey interviews (with comments and rejoinder). *Journal of the American Statistical Association*, 85(March), 232–253.

Thurstone, L.L. (1927) A law of comparative judgment. *Psychological Review*, 34(4), 273.
Vannette, D. (2017) Using attention checks in your surveys may harm data quality. Available from: https://www.qualtrics.com/blog/using-attention-checks-in-your-surveys-may-harm-data-quality/
Yu, P.L. & Chan, L.K. (2001) Bayesian analysis of wandering vector models for displaying ranking data. *Statistica Sinica*, 11, 445–461.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Yu, Q. & Quinn, K.M. (2022) A multidimensional pairwise comparison model for heterogeneous perceptions with an application to modelling the perceived truthfulness of public statements on COVID-19. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 185(3), 1049–1073. Available from: https://doi.org/10.1111/rssa.12810

APPENDIX A. MARKOV CHAIN MONTE CARLO ALGORITHM

In this section, we detail the samplers for the two versions of the new model. The model fitting algorithms can be found in the `MCMCpack` R package which is available at https://cran.r-project.org/.

A.1 Sampler for the first version of the new model

The sampler for the first version of the new model consists of a Gibbs sampler component and a random walk Metropolis–Hastings (MH) sampler component. We use the Gibbs sampler to sample $\boldsymbol{\theta}_j$'s and the augmented parameters, $y^*_{ijj'}$'s. We use the random walk MH sampler to sample γ_i's.

A.1.1 Sample $y^*_{ijj'}$

We use the data augmentation method in Bayesian statistics to replace the binary choice data points with continuous values (Albert & Chib, 1993). A binary choice data point has the following Bernoulli distribution:

$$y_{ijj'} \sim \mathcal{B}ernoulli(p_{ijj'})$$
$$p_{ijj'} = \Phi_1(\boldsymbol{\theta}_j \cdot \mathbf{g}(\gamma_i) - \boldsymbol{\theta}_{j'} \cdot \mathbf{g}(\gamma_i))$$

We define a continuous latent attribute difference, $y^*_{ijj'}$, to correspond every binary data point, $y_{ijj'}$. We use $\varepsilon_{ijj'} \sim N(0,1)$ to denote an i.i.d error term. Therefore, $y_{ijj'} = 1$ is equivalent to a positive attribute difference between object j and object j' for respondent i: $y^*_{ijj'} = \boldsymbol{\theta}_j \cdot \mathbf{g}(\gamma_i) - \boldsymbol{\theta}_{j'} \cdot \mathbf{g}(\gamma_i) + \varepsilon_{ijj'} > 0$. Likewise, $y_{ijj'} = 0$ is equivalent to a negative attribute difference between object j and object j' for respondent i: $y^*_{ijj'} = \boldsymbol{\theta}_j \cdot \mathbf{g}(\gamma_i) - \boldsymbol{\theta}_{j'} \cdot \mathbf{g}(\gamma_i) + \varepsilon_{ijj'} < 0$. Without loss of generality, we impose a truncated standard normal distribution on $y^*_{ijj'}$. The sign of $y^*_{ijj'}$ must be equal to the sign of the corresponding $y_{ijj'} - \frac{1}{2}$. Given the current values of $\boldsymbol{\theta}_j$, $\boldsymbol{\theta}'_j$ and γ_i, we can sample $y^*_{ijj'}$ from the following distribution:

$$y^*_{ijj'} \sim \begin{cases} \mathcal{N}(\boldsymbol{\theta}_j \cdot \mathbf{g}(\gamma_i) - \boldsymbol{\theta}_{j'} \cdot \mathbf{g}(\gamma_i), 1)\mathbb{I}(y^*_{ijj'} > 0), & \text{if } y_{ijj'} = 1 \\ \mathcal{N}(\boldsymbol{\theta}_j \cdot \mathbf{g}(\gamma_i) - \boldsymbol{\theta}_{j'} \cdot \mathbf{g}(\gamma_i), 1)\mathbb{I}(y^*_{ijj'} < 0), & \text{if } y_{ijj'} = 0 \end{cases}$$

where $\mathcal{N}(\boldsymbol{\theta}_j \cdot \mathbf{g}(\gamma_i) - \boldsymbol{\theta}_{j'} \cdot \mathbf{g}(\gamma_i), 1)\mathbb{I}(y^*_{ijj'} > 0)$ is a univariate truncated normal distribution, which only takes positive values. Similarly, $\mathcal{N}(\boldsymbol{\theta}_j \cdot \mathbf{g}(\gamma_i) - \boldsymbol{\theta}_{j'} \cdot \mathbf{g}(\gamma_i), 1)\mathbb{I}(y^*_{ijj'} < 0)$ is a univariate truncated normal distribution, which only takes negative values.

A.1.2 Sample θ_j

For sampling the values of $\boldsymbol{\theta}_j$, we need to do careful bookkeeping of all the pairwise comparison tasks that involve object j. Let us use M_j to denote the number of all the unique comparison tasks involving object j. Accordingly, we need to record the sign of $\boldsymbol{\theta}_j$, the counterpart object attribute $\boldsymbol{\theta}_{j'}$, the respondent attribute γ_i, and the augmented parameter $y^*_{ijj'}$ or $y^*_{ij'j}$ (depending on which side object j shows in this task) in each one of the M_j tasks.

We use the rows of matrix $\tilde{\boldsymbol{\Theta}}$ to store the counterpart object attribute $\boldsymbol{\theta}_{j'}$'s in all the comparison tasks involving object j. We use vector $\tilde{\boldsymbol{\gamma}}_j$ to store the respondent attribute γ_i's in all the comparison tasks involving object j. We use vector $\tilde{\boldsymbol{y}}^*_j$ to store the augmented parameter y^*'s in all the comparison tasks involving object j. When object j shows up in a comparison task, it either shows as the left-side choice or right-side choice. For a unique comparison task, we denote the sign of object j as $+1$ if it is the left-side option, or as -1 if it is the right-side option. We use vector $\boldsymbol{s}_j$ to store the signs of object j in all the comparison tasks involving object j. These four containers are filled in a specific order so that the m'th element or row of the four containers correspond to the parameters for the m'th comparison task involving object j.

To illustrate the relationship between $\boldsymbol{\theta}_j$, $\tilde{\boldsymbol{\Theta}}_j$, $\tilde{\boldsymbol{\gamma}}_j$, and $\tilde{\boldsymbol{y}}^*_j$, we write out the equation representing the m'th comparison involving object j.

$$\boldsymbol{\theta}_j \cdot (\boldsymbol{s}_j[m] \times \mathbf{g}(\tilde{\boldsymbol{\gamma}}_j[m])) + \text{error term} = \tilde{\boldsymbol{y}}^*_j[m] + (\boldsymbol{s}_j[m] \times (\tilde{\boldsymbol{\Theta}}_j[m] \cdot \mathbf{g}(\tilde{\boldsymbol{\gamma}}_j[m])))$$

where $[m]$ indicates the m's element of a vector or the m's row of a matrix, and the error term has an i.i.d standard normal distribution.

We can write a similar equation for every comparison task involving object j. The left-hand side of the equation is a dot product of $\boldsymbol{\theta}_j$ and another vector, and the right-hand side of the equation is a scalar. For $m = 1, 2, \ldots, M_j$, we repeat the same algebraic manipulation in the above equation. We compute $\boldsymbol{s}_j[m] \times \mathbf{g}(\tilde{\boldsymbol{\gamma}}_j[m])$, and store the resulting vector in the m'th row of matrix $\boldsymbol{X}_j$. We compute $\tilde{\boldsymbol{y}}^*_j[m] + (\boldsymbol{s}_j[m] \times (\tilde{\boldsymbol{\Theta}}_j[m] \cdot \mathbf{g}(\tilde{\boldsymbol{\gamma}}_j[m])))$, and store the resulting value in the m'th element of vector $\boldsymbol{z}_j$. Then we can express the distribution of $\boldsymbol{z}_j$ with the multidimensional normal distribution below.

$$\underbrace{\boldsymbol{z}_j}_{M_j \times 1 \text{ vector}} \sim \mathcal{N}_{M_j}\left(\underbrace{\boldsymbol{X}_j}_{M_j \times 2 \text{ matrix}} \times \underbrace{\boldsymbol{\theta}_j}_{2 \times 1 \text{ vector}}, \mathbf{I}_{M_j}\right)$$

$\boldsymbol{\theta}_j$ has a semi-conjugate bivariate normal prior distribution.

$$\boldsymbol{\theta}_j \sim \mathcal{N}_2(\mathbf{0}, \mathbf{I}_2)$$

Then, we are able to derive the conditional posterior of $\boldsymbol{\theta}_j$ as follows:

$$\boldsymbol{\theta}_j|\boldsymbol{X}_j, \boldsymbol{z}_j \sim \mathcal{N}_2\left((\boldsymbol{X}_j^T\boldsymbol{X}_j + \mathbf{I}_2)^{-1}\boldsymbol{X}_j^T\boldsymbol{z}_j, \quad (\boldsymbol{X}_j^T\boldsymbol{X}_j + \mathbf{I}_2)^{-1}\right)$$

A.1.3 Sample γ_i

For sampling the values of γ_i, we need to do careful bookkeeping of all the pairwise comparison tasks that involve respondent i. Let us use M_i to denote the number of all the unique comparison tasks involving respondent i. Accordingly, we need to record the left-side object attribute $\boldsymbol{\theta}_j$, the right-side object attribute $\boldsymbol{\theta}_{j'}$, and the augmented parameter $y^*_{ijj'}$ in each one of the M_i tasks.

We use the rows of matrix $\tilde{\boldsymbol{\Theta}}_i$ to store the left-side object attribute $\boldsymbol{\theta}_j$'s in all the comparison tasks involving respondent i. We use the rows of matrix $\tilde{\boldsymbol{\Theta}}'_i$ to store the right-side object attribute $\boldsymbol{\theta}_{j'}$'s in all the comparison tasks involving respondent i. We use vector $\tilde{\boldsymbol{y}}^*_i$ to store the augmented parameter y^*'s in all the comparison tasks involving respondent i. These three containers are filled in a specific order so that the m'th element or row of the three containers correspond to the parameters for the m'th comparison task involving respondent i.

Given the current values of $\gamma_i^{(t)}$, $\tilde{\boldsymbol{\Theta}}_i$, and $\tilde{\boldsymbol{\Theta}}'_i$, the density function of $\tilde{\boldsymbol{y}}^*_i$ is the product of M_i normal distribution densities:

$$\mathcal{L}\left(\tilde{\boldsymbol{y}}^*_i|\gamma_i^{(t)}, \tilde{\boldsymbol{\Theta}}_i, \tilde{\boldsymbol{\Theta}}'_i\right) = \prod_{m=1}^{M_i}\phi_1\left(\tilde{\boldsymbol{y}}^*_i[m]; \quad \tilde{\boldsymbol{\Theta}}_i[m]\cdot\mathbf{g}\left(\gamma_i^{(t)}\right) - \tilde{\boldsymbol{\Theta}}'_i[m]\cdot\mathbf{g}\left(\gamma_i^{(t)}\right), 1\right)$$

where $[m]$ indicates the m'th element of a vector or the m'th row of a matrix, and $\phi_1(\cdot;\, \mu,\, \sigma^2)$ is the PDF of a univariate normal distribution with mean μ and variance σ^2.

We use a random walk MH sampler to sample γ_i, and we sample each γ_i separately for $i = 1, 2, \ldots, I$. We generate a random walk step, τ, from a uniform distribution, $\tau \sim \mathcal{U}nif(-\delta, \delta)$. δ is the positive tuning parameter that determines the accepting rate of the random walk MH sampler. (Chib & Greenberg, 1995) The proposed new value of γ_i is $\gamma_i^{(t+1)} = \gamma_i^{(t)} + \tau$. We plug $\gamma_i^{(t+1)}$ in the density function, and get $\mathcal{L}(\tilde{\boldsymbol{y}}^*_i|\gamma_i^{(t+1)}, \tilde{\boldsymbol{\Theta}}_i, \tilde{\boldsymbol{\Theta}}'_i)$.

Due to the uniform prior on γ_i, the acceptance ratio, r, is determined only by the ratio of the likelihoods.

$$r = \min\left(1, \frac{\mathcal{L}\left(\tilde{\boldsymbol{y}}^*_i|\gamma_i^{(t+1)}, \tilde{\boldsymbol{\Theta}}_i, \tilde{\boldsymbol{\Theta}}'_i\right)}{\mathcal{L}\left(\tilde{\boldsymbol{y}}^*_i|\gamma_i^{(t)}, \tilde{\boldsymbol{\Theta}}_i, \tilde{\boldsymbol{\Theta}}'_i\right)}\right)$$

With probability r, we accept the proposed new $\gamma_i^{(t+1)}$, and with probability $1 - r$, we reject it.

A.2 Sampler for the second version of the new model

The sampler for the second version of the new model shares the same steps for sampling $y^*_{ijj'}$ and $\boldsymbol{\theta}_j$ in the first version. We only introduce the rest of the steps in the sampler for the second version of the new model, given the current values of $y^*_{ijj'}$'s and $\boldsymbol{\theta}_j$'s. We assume a Dirichlet process prior on γ_i. Before specifying the sampler for γ_i, we compare two approaches to implement a Dirichlet process Mixture model: the collapsed sampler and the blocked Gibbs sampler (Müller & Rodriguez, 2013).

The collapsed sampler approach analytically computes the probability for assigning a unit to a cluster by integrating out the parameters characterizing each cluster (Escobar, 1994; Escobar & West, 1995; Ferguson, 1973; MacEachern, 1994; MacEachern & Müller, 1998; Neal,

2000). This approach works well with conjugate priors, and cleverly uses the integral trick to account for infinite values of the cluster parameters when deciding a cluster assignment probability. The collapsed sampler has wide applications in various fields. The limitation of the collapsed sampler lies in the relative difficulty for it to work with non-conjugate priors.

The blocked Gibbs sampler has its theoretical foundation in the stick-breaking process reparameterization of the Dirichlet process (Ishwaran & James, 2001; Ishwaran & Zarepour, 2000; Sethuraman, 1994). The blocked Gibbs sampler further simplifies the sampling procedure by assuming a finite number of candidate clusters to start with (Müller & Rodriguez, 2013). There are other augmented variables in the blocked Gibbs sampler to facilitate large clusters to grow larger and small clusters to disappear. Even if we assume a large number of finite candidate clusters at the beginning, the block Gibbs sampler will eventually converge to a small number of clusters as the MCMC mixes. Therefore, the block Gibbs sampler represents a close and efficient approximation to the original Dirichlet process with infinite candidate clusters.

Due to the non-conjugate prior employed on γ_i, we use the block Gibbs sampler for sampling γ_i in the second version of the new model. In this subsection, we first specify the Dirichlet process prior on γ_i. Then we introduce the sampling steps for the Dirichlet process part of the second version of the new model.

We assume a finite maximum number of clusters K. We denote each cluster membership of respondent i as L_i, $L_i \in \{1, 2, \ldots, K\}$. Cluster k is characterized by parameter, γ_k. In contrast to the first version of the new model where each respondent has a unique γ_i, different respondents may share the same γ_k if they are in the same cluster k in the second version. All the γ_k's have the Dirichlet process prior.

$$\gamma_k \overset{iid}{\sim} G \quad k = 1, \ldots, K$$
$$G \sim \mathcal{DP}(\alpha G_0)$$

where $\alpha \in \mathbb{R}_+$ is a concentration parameter and G_0 is the centring distribution, which is specified as $\mathcal{U}nif(0, \frac{1}{2}\pi)$.

Without considering any density function, we devise an augmented cluster weight parameter, ω_k. The purpose of ω_k's is to induce sparsity in clustering, so that large clusters tend to grow larger and small clusters tend to disappear. The prior for the vector $\boldsymbol{\omega}$ is a generalized Dirichlet distribution (Connor & Mosimann, 1969; Ishwaran & James, 2001). ω_k is generated from a stick-breaking process, for $k = 1, 2, \ldots, K$, and is only determined by the current sizes of all the clusters.

To express the density function of the augmented parameters, $y^*_{ijj'}$'s, associated with respondent i, conditioning on respondent i being in cluster k, we need to use the notations elaborated in the last section, $\tilde{\boldsymbol{\Theta}}_i$, $\tilde{\boldsymbol{\Theta}}'_i$, and $\tilde{\boldsymbol{y}}^*_i$. Given respondent i being in cluster k with γ_k, the conditional density function of $\tilde{\boldsymbol{y}}^*_i$ is the product of M_i normal distribution densities:

$$\mathcal{L}\left(\tilde{\boldsymbol{y}}^*_i | \gamma_k, \tilde{\boldsymbol{\Theta}}_i, \tilde{\boldsymbol{\Theta}}'_i\right) = \prod_{m=1}^{M_i} \phi_1\left(\tilde{\boldsymbol{y}}^*_i[m]; \quad \tilde{\boldsymbol{\Theta}}_i[m] \cdot \mathbf{g}(\gamma_i) - \tilde{\boldsymbol{\Theta}}'_i[m] \cdot \mathbf{g}(\gamma_k), 1\right)$$

Both the cluster weight, ω_k, and the conditional density of respondent i's augmented parameters, $\tilde{\boldsymbol{y}}^*_i$, given respondent i being in cluster k, contribute to the probability of assigning respondent

i to cluster k. We denote the probability of assigning respondent i to cluster k as q_{ik}. For $k = 1, 2, \ldots, K$, we compute q_{ik} as follows:

$$q_{ik} \propto \omega_k \mathcal{L}\left(\tilde{\boldsymbol{y}}_i^* | \gamma_k, \tilde{\boldsymbol{\Theta}}_i, \tilde{\boldsymbol{\Theta}}_i'\right)$$
$$\sum_{k=1}^{K} q_{ik} = 1$$

The cluster label for respondent i has a categorical distribution:

$$L_i \sim Categorical(q_{i1}, q_{i2}, \ldots, q_{iK})$$

We can supply a fixed value for the precision parameter, α. Or we can treat α as a parameter to estimate based on the data. In the latter case, we put a conjugate Gamma prior on α with shape a and rate b:

$$\alpha \sim \mathcal{G}amma(a, b)$$

In the rest of this subsection, we demonstrate the steps to sample the parameters above.

A.2.1 Sample γ_k

Given the current cluster memberships of all the respondents, we update each cluster's γ_k with either a mini random walk Metropolis–Hastings sampler or a simple draw from the prior. If cluster k is empty, then we don't have any empirical data for updating γ_k. We simply draw a new γ_k from $\mathcal{U}nif(0, \frac{\pi}{2})$. If cluster k has members, we treat these respondents and their associated $\tilde{\boldsymbol{\Theta}}_i$, $\tilde{\boldsymbol{\Theta}}_i'$, and $\tilde{\boldsymbol{y}}_i^*$ as belonging to cluster k. Then we use a random walk MH sampler to update γ_k. Theoretically, we can use a one-step MH sampler for each cluster, and the MCMC should eventually traverse to the mode of each γ_k. In order to improve the efficiency of MCMC, we do multiple steps of MH sampler and update γ_k with the last-step value. We need to specify the iteration number and tuning parameters for these mini MH samplers.

In each iteration within a mini MH sampler, we do the following steps. Given the current value of $\gamma_k^{(t)}$, $\{\tilde{\boldsymbol{\Theta}}_i\}_{i:L_i=k}$, and $\{\tilde{\boldsymbol{\Theta}}_i'\}_{i:L_i=k}$, the conditional density of $\{\tilde{\boldsymbol{y}}_i^*\}_{i:L_i=k}$ is the product of multiple normal distribution densities:

$$\mathcal{L}\left(\{\tilde{\boldsymbol{y}}_i^*\}_{i:L_i=k} | \gamma_k^{(t)}, \{\tilde{\boldsymbol{\Theta}}_i\}_{i:L_i=k}, \{\tilde{\boldsymbol{\Theta}}_i'\}_{i:L_i=k}\right)$$
$$= \prod_{i:L_i=k} \prod_{m=1}^{M_i} \phi_1\left(\tilde{\boldsymbol{y}}_i^*[m]; \quad \tilde{\boldsymbol{\Theta}}_i[m] \cdot \mathbf{g}\left(\gamma_k^{(t)}\right) - \tilde{\boldsymbol{\Theta}}_i'[m] \cdot \mathbf{g}\left(\gamma_k^{(t)}\right), 1\right)$$

where $[m]$ indicates the m'th element of a vector or the m'th row of a matrix, and $\phi_1(\cdot)$ is the PDF of a univariate normal distribution.

We generate a random walk step, τ, from a uniform distribution, $\tau \sim \mathcal{U}nif(-\delta, \delta)$. δ is the positive tuning parameter that determines the accepting rate of the mini random walk MH sampler. The proposed new value of γ_k is $\gamma_k^{(t+1)} = \gamma_k^{(t)} + \tau$. We plug $\gamma_k^{(t+1)}$ in the density function, and get $\mathcal{L}(\{\tilde{\boldsymbol{y}}_i^*\}_{i:L_i=k} | \gamma_k^{(t+1)}, \{\tilde{\boldsymbol{\Theta}}_i\}_{i:L_i=k}, \{\tilde{\boldsymbol{\Theta}}_i'\}_{i:L_i=k})$.

Due to the uniform prior on γ_k, the acceptance ratio, r, is determined only by the ratio of the density functions.

$$r = \min\left(1, \frac{\mathcal{L}(\{\tilde{\mathbf{y}}_i^*\}_{i:L_i=k} | \gamma_k^{(t+1)}, \{\tilde{\mathbf{\Theta}}_i\}_{i:L_i=k}, \{\tilde{\mathbf{\Theta}}_i'\}_{i:L_i=k})}{\mathcal{L}(\{\tilde{\mathbf{y}}_i^*\}_{i:L_i=k} | \gamma_k^{(t)}, \{\tilde{\mathbf{\Theta}}_i\}_{i:L_i=k}, \{\tilde{\mathbf{\Theta}}_i'\}_{i:L_i=k})}\right)$$

With probability r, we accept the proposed new $\gamma_k^{(t+1)}$, and with probability $1-r$, we reject it. We store the last-step value $\gamma_k^{(T)}$, and update the old γ_k with $\gamma_k^{(T)}$. We repeat this process for each cluster, until we finish updating all the γ_k's.

A.2.2 Sample ω_k

Given the current respondents' cluster memberships and each cluster's size, we use a stick-breaking process to update ω_k's. We denote the size of cluster k as ζ_k. To generate ω_k, we need to introduce auxiliary parameters, V_k, for $k = 1, 2, \ldots, K-1$. Given the current cluster sizes, ζ_k's, we first generate the auxiliary parameters, V_k as below.

$$V_k \sim \mathcal{B}eta\left(1 + \zeta_k, \alpha + \sum_{l=k+1}^{K} \zeta_l\right), \text{ for } k = 1, 2, \ldots, K-1$$

Then we update ω_k's according to the following formula:

$$\begin{aligned}
\omega_1 &= V_1 \\
\omega_k &= V_k \prod_{l=1}^{k-1}(1 - V_l), \text{ for } k = 2, \ldots, K-1 \\
\omega_K &= \prod_{l=1}^{K-1}(1 - V_l) = 1 - \sum_{l=1}^{K-1}\omega_l
\end{aligned}$$

A.2.3 Sample L_i

Given the current value of γ_k for cluster k, we compute the conditional density, $\mathcal{L}(\tilde{\mathbf{y}}_i^* | \gamma_k, \tilde{\mathbf{\Theta}}_i, \tilde{\mathbf{\Theta}}_i')$, for respondent i to be in cluster k. We then take the product of ω_k and $\mathcal{L}(\tilde{\mathbf{y}}_i^* | \gamma_k, \tilde{\mathbf{\Theta}}_i, \tilde{\mathbf{\Theta}}_i')$, and use it to form the categorical distribution below to draw the new cluster label, L_i, for respondent i, from the discrete cluster label set, $\{1, 2, \ldots, K\}$.

$$\begin{aligned}
&L_i \sim \mathcal{C}ategorical(q_{i1}, q_{i2}, \ldots, q_{iK}) \\
&q_{ik} \propto \omega_k \mathcal{L}(\tilde{\mathbf{y}}_i^* | \gamma_k, \tilde{\mathbf{\Theta}}_i, \tilde{\mathbf{\Theta}}_i') \\
&\sum_{k=1}^{K} q_{ik} = 1
\end{aligned}$$

A.2.4 Sample α

The vector $\boldsymbol{\omega}$ has a generalized Dirichlet distribution prior, and the concentration parameter α is a parameter in this prior. The conditional distribution of α, given the current $\boldsymbol{\omega}$, has the kernel of a Gamma distribution: $f(\alpha|\boldsymbol{\omega}) \propto \alpha^{K-1}\omega_K^{\alpha} = \alpha^{K-1}\exp(-(-\alpha \log \omega_K))$ (Ishwaran & Zarepour, 2000). Given the conjugate Gamma prior on α, $\alpha \sim \mathcal{G}amma(a, b)$, we can express the conditional posterior for α as follows:

$$\alpha|\boldsymbol{\omega} \sim \mathcal{G}amma(a + K - 1, b - \log \omega_K)$$

APPENDIX B. ADDITIONAL INFORMATION ON THE SURVEY

This survey was judged exempt from review by our university's IRB (study ID HUM00184241).

After a respondent provided their informed consent to continue with the survey, a short training page was provided to the respondent. This training page made it clear that the questions about the relative truthfulness of pairs of COVID-19 statements were eliciting the respondent's belief about which statement was more truthful when it was stated. The key language here was: 'Factual statements can be placed on a line. At one extreme end of the line are statements that are completely truthful and accurate. At the other extreme end are statements that are intentionally false. Between these two extremes we find statements that contain elements of truth and falsity and/or half-truths. For each task, you will see two statements on the coronavirus pandemic. Your task is to read both and to select the statement that you believe was **more truthful when it was stated**'.

After this brief training, each respondent was given a single attention check question that provided the respondent with two COVID-19 statements and asked them to select both statements. The text of the question was: 'The following are two statements about the coronavirus pandemic. These statements were made between late February and early May, 2020. We are interested in which statement you believe was more truthful when it was made. However, for this question, we care more about whether you are paying attention. Please choose both the first and second statements to indicate you are paying attention'. Recent work on attention checks in online surveys suggests that eliminating respondents who fail attention checks may introduce demographic bias (Vannette, 2017). Consequently, we do not exclude respondents who fail this check. The purpose of including this attention check is solely to encourage respondents to read the following response prompts carefully.

As a robustness check, we subset the data down to only those respondents who passed the attention check. Among the 2,621 usable respondents, 1,136 passed the attention check. We then replicate the main analysis with the two-dimensional Dirichlet process model using this subset of data. We find that the results are qualitatively similar to the full data results. We report these robustness check results in the supplemental information document.