Journal of Educational and Behavioral Statistics
http://jebs.aera.net

Beta Regression Finite Mixture Models of Polarization and Priming

Michael Smithson, Edgar C. Merkle and Jay Verkuilen

JOURNAL OF EDUCATIONAL AND BEHAVIORAL STATISTICS 2011 36: 804
originally published online 20 June 2011
DOI: 10.3102/1076998610396893

Journal of Educational and Behavioral Statistics
December 2011, Vol. 36, No. 6, pp. 804–831
DOI: 10.3102/1076998610396893

Beta Regression Finite Mixture Models of Polarization and Priming

Michael Smithson
The Australian National University

Edgar C. Merkle
Wichita State University

Jay Verkuilen
Graduate Center, City University of New York

This paper describes the application of finite-mixture general linear models based on the beta distribution to modeling response styles, polarization, anchoring, and priming effects in probability judgments. These models, in turn, enhance our capacity for explicitly testing models and theories regarding the aforementioned phenomena. The mixture model approach is superior in this regard to popular methods such as extremity scores, due to its incorporation of three submodels (location, dispersion, and relative composition), each of which can diagnose specific kinds of polarization and related effects. Three examples are elucidated using real data sets.

Keywords: *beta distribution; mixture model; polarization; priming; anchoring*

1. Introduction

Most of the research on probability judgments and other doubly bounded scales has used traditional analysis of variance (ANOVA) and linear regression to model responses, ignoring the scale bounds, assuming homogeneity of variance, and focusing exclusively on modeling the mean response. Likewise, research on attitude polarization, anchoring, and priming effects has been hampered by a lack of appropriate and useful analytical techniques. For the most part, experimental studies on these topics use normal-theory linear regression (predominantly ANOVA) to track mean responses. Data from studies on these topics can be multimodal with severe skews in both directions, heteroscedastic, and have floor or ceiling effects. Traditional methods such as normal-theory linear regression are problematic, first because they ignore scale bounds, and second because the assumptions required by such methods preclude investigating several distinct and relevant phenomena.

In addition to attitudinal extremity and polarization, these problems extend to cognitive topics such as priming, anchoring, additivity of probability judgments, and probability weighting.

This paper describes the use of finite-mixture generalized linear models (GLMs) based on the beta distribution for modeling response styles and related phenomena in probability judgments. These models, in turn, enhance our capacity for explicitly testing models and theories regarding the aforementioned phenomena. We will focus primarily on polarization, anchoring, and priming effects.

2. Extremity and Polarization

Polarization and extreme-response phenomena are of interest primarily in social and organizational psychology, especially in research on attitudes and norms. These phenomena typically arise in settings where there are sharply divided groups and/or strong pressures toward conformity. However, they also can arise in cognitive psychological research, particularly in processing ambiguous stimuli. Extreme responses refer to either strong endorsement of or strong opposition to an attitude or belief. Polarization refers to strong disagreement between individuals or groups on an issue, such that some people strongly endorse and others strongly oppose the attitude or belief concerned, with few taking a middle position. Extreme-response phenomena thereby include polarization as a special case. Because an indication of strength of endorsement is required to identify whether polarization or extreme response has occurred, studies of polarization utilize response scales designed to measure this strength.

A widely used approach to studying polarization is extremity scores, that is, absolute deviations from scale midpoints (e.g., Brauer, Judd, & Gliner, 1995; Downing, Judd, & Brauer, 1992). Extremity scores usually are modeled via conventional GLMs, although it is debatable whether normal-theory GLMs are appropriate for them. However, extremity scores are potentially misleading on the following grounds:

1. Because extremity scores ignore the sign of the deviation, it is possible for extremism to exist without polarization. In fact, extremity scores are not always capable of distinguishing between a unimodal strongly skewed distribution (as in a strongly consensual sample of extremists) from a bimodal bi-skewed distribution (as in a strongly polarized sample of extremists). If the bimodal distribution were symmetrical, then extremity scores would superimpose the left-side component distribution onto its right-side twin, rendering this case indistinguishable from the case where all of the data were concentrated in one or the other component distribution.
2. The relative size of polarized groups may change without mean extremity changing when the groups are equidistant from the center of the scale. In this case, extremity scores are incapable of distinguishing between a tiny minority–large majority polarization and an equal-sized groups polarization.

3. The overlap between polarized groups may change without mean extremity changing, because the means may be unaffected by changes in dispersion. Mean-response models of extremity scores cannot distinguish between two widely dispersed overlapping distributions and two tightly clumped nonoverlapping distributions with the same means.

Thus, polarization cannot be completely modeled using extremity scores because simple extremism ignores essential specifics of polarization.

A second approach to studying polarization posits two (or more) subpopulations that are distinguishable in their distributions on an attitude measure. Latent class (Heinen, 1996) and taxometric (Waller & Meehl, 1998) techniques are examples of this approach. In the past two decades, several models that combine latent class and latent trait models have been developed to allow a distinct latent trait model to apply within each latent class (see Fieuws, Spiessens, & Draney, 2004). The approach we adopt here belongs to the same class of mixture models. However, hybrid latent class–latent trait models are oriented primarily toward the assumption that latent categories (or taxa) exist. Unlike taxa, polarization is not an all-or-nothing phenomenon and therefore requires models that predict when it will appear and disappear or wax and wane. Our approach permits polarization to manifest itself in varying degrees. The distance between polarized clumps, their degree of overlap, and their relative sizes all may vary and these can be modeled using our framework.

Our approach starts by considering a finite mixture-distribution that models polarization effects as influencing component distribution means, precision, and relative composition. The model assumes two or more subpopulations, each with its own component distribution. Given a collection of $Y_1, \ldots, Y_n$ independent identically distributed random variables, the probability density function (pdf) of Y_i is expressible as a weighted sum of two or more component pdfs:

$$\sum_j \gamma_{ji} f_{ji}(y), \tag{1}$$

for $j = 1, \ldots, J$, where $0 \leq \gamma_{ji} \leq 1$ and $\sum_j \gamma_{ji} = 1$.

If Y also is bounded below and above (doubly bounded), then we require a distribution whose support is restricted to the range of Y. Without loss of generality, we shall assume from here on that the support is the [0,1] interval. In some applications, we shall assume each $f_{ji}(y)$ is a Beta (ω_j, τ_j) pdf. We reparameterize these component distributions in terms of a location (mean), $\mu_j = \omega_j/(\omega_j + \tau_j)$ and precision parameter $\phi_j = \omega_j + \tau_j$ (for further details, see Smithson & Verkuilen, 2006). Note that the term "precision" is used differently here from its conventional meaning as the reciprocal of the variance. This precision parameter is independent of the mean, whereas the variance of the beta distribution, $\sigma^2 = \mu(1 - \mu)/(\phi + 1)$, is not.

The resulting GLM has three submodels, whereby we can individually examine effects of predictor variables on the location, precision, and relative composition parameters. The *location submodel* is

$$g(\mu_{ji}) = \sum_{k} \beta_{jk} X_{jki}, \tag{2}$$

for $j = 1, 2, ..., J - 1$ and for $k = 0, 1, ..., K$ where the link function is the logit $g(v) = \log(v/(1 - v))$, the X_{jki} are predictors and the β_{jk} are coefficients. Thus, this submodel predicts a change of β_{jk} in the logit of μ_{ji} for every unit change in X_{jki}. The *dispersion submodel* is

$$h(\phi_{ji}) = \sum_{m} -\delta_{jm} W_{jmi}, \tag{3}$$

for $m = 0, 1, \ldots, M$ where $h(v) = \log(v)$. This submodel is related to dispersion as well as precision because of the negative sign given to the δ_{jm} coefficients, so that larger values predict greater dispersion (lower precision). The *relative composition submodel* (predicting the relative size of the component distributions) is

$$\gamma_{ji} = \frac{\exp\left(\sum_{p} \theta_{jp} Z_{jpi}\right)}{1 + \sum_{k=1}^{J-1} \exp\left(\sum_{p} \theta_{kp} Z_{kpi}\right)}, \tag{4}$$

for $j = 1, \ldots, J - 1$ and $p = 0, 1, \ldots, P$. The Jth component γ_{Ji} is defined by

$$\gamma_{Ji} = 1 - \sum_{j=1}^{J-1} \gamma_{ji}.$$

This submodel is linearizable via the inverse transformation

$$\log\left(\gamma_{ji}/\gamma_{Ji}\right) = \sum\nolimits_{p} \theta_{jp}\, Z_{jpi}. \tag{5}$$

All three submodels may be simultaneously estimated using the standard maximum likelihood approach.

Mixture models can be unidentified, so the question of identifiability is reasonable to raise here. Although it is not possible to provide a definitive answer, we ran a simulation in Mathematica v.7 for a two-component model, varying the beta distribution parameters to simulate three conditions:

1. Fairly precise distributions that may overlap, with ω_j and τ_j, given uniform distributions over the [1,25] interval;
2. One fairly imprecise distribution and a precise distribution, with ω_1 given a uniform distribution over [0.1,1], and ω_2 and τ_j given uniform distributions over [1,10]; and

3. Two fairly imprecise distributions that may overlap, with ω_j given a uniform distributions over [0.1,1], and τ_j given uniform distributions over [1,10].

Each condition was run 5,000 times and the rank of the Jacobian matrix was computed for each run. Of the 15,000 runs, 14,758 (98.4%) produced a full-rank Jacobian, suggesting that this model is identified a large portion of the time under realistic conditions. Of course, this demonstration does not obviate the need to check whether a model is identified in a particular application.

As Verkuilen and Smithson (in press) observe, beta regression GLMs pose problems for model diagnostics that are open questions in the GLMM literature, and a complete discussion of these is beyond the scope of this paper. Model comparison under maximum likelihood estimation is reasonably straightforward because the beta is a member of the exponential family. Model evaluation and checking entail three issues: How accurately the model "predicts" the data, the influence of individual data points, and how appropriately the model assigns cases to component distributions. The fit between the model and the data can be evaluated by assessing how well the mean and variance structures are reproduced and via simulations from the posterior predictive density. Influence poses some difficulties, because even a one-component beta GLM lacks an appropriate residual or deviance. The alternatives proposed in the recent literature for assessing influence are reviewed in Verkuilen and Smithson.

In many applications of the kind we have in mind here, case membership in component distributions is unobserved so there is no way to assess how accurately the model recovers the assignment of cases to component distributions. Nevertheless, if none of the component distributions is degenerated, then the model may be evaluated for how well separated the component posterior predictive densities are at each data point. We illustrate such procedures in our examples.

We can have models in which covariates predict one or more means of the $f_{ji}(y)$, the mixture parameters, or the precision parameters of the $f_{ji}(y)$. Thus, there are three distinguishable, "pure" kinds of polarization phenomena:

1. Location drift: Only the component distribution means are predicted by covariates. Polarization changes only as a function of the distance between the means of the component distributions, whose precision parameters and relative sizes remain constant.
2. Dispersion drift: Only the component distribution precision parameters are predicted by covariates. Polarization changes as a function of the overlap between the component distributions, whose means and relative sizes remain constant.
3. Composition shift: Only the relative composition parameters are predicted by covariates. Polarization changes only as a function of the relative sizes of the component distributions, whose means and variances remain constant.

It may be somewhat unusual to find pure instances of any of these, but the crucial point is that the approach is capable of distinguishing among the three manifestations of polarization and separate effects of independent variables on each of them, via the three submodels developed earlier. We shall see that this separability enables testing hypotheses that otherwise would be misspecified and identifying effects that otherwise would be obscured.

3. Specified Anchors

Probability judgment tasks require judges to assign (usually numerical) estimates of probabilities of events. Judges are said to "anchor" their estimates on a particular value if their initial estimates tend to be close to that value and are shifted by new evidence to a lesser extent than would be the case for a Bayesian agent. Some research on probability judgments investigates whether judges can be "primed" to focus on a specific anchor. When the location of an anchor can be specified a priori, a reasonable choice of mixture model has one component distribution's location parameter fixed at the anchor location. In the literature on probability judgments, partition priming presents one example of specifiable anchors, where subjective prior probabilities are centered on $1/K$ by judges being primed to believe that there are K possible events (see, e.g., Fox & Rottenstreich, 2003). Where there is a normatively correct partition, we say that the anchor is "normatively specified." Normatively specified anchors from research on probability judgments include:

1. Anchoring on $1/K$ when there is a correct K-fold partition;
2. Additivity when probabilities are required to sum to 1 across a collection of events;
3. "Correct" conditional or compound probabilities (e.g., according to the rules of probability theory or Bayesian updating); and
4. Conjugacy of lower and upper probabilities (i.e., lower $P(A) = 1 -$ upper $P(NotA)$, where lower $P(A)$ and upper $P(A)$ are an interval containing $P(A)$).

Examples of specified anchors that are not normative include:

1. Anchoring on $1/K$, given an arbitrary (or incorrect) K-fold partition;
2. Additivity for probabilities of events that do not form an exhaustive, mutually exclusive collection of events; and
3. The use of 1/2 as a probability assignment for signifying complete ignorance of the likelihood of an event, regardless of how many events there are in the partition.

The fact that these anchors are specified pointwise has implications for constructing a mixture model to test them. First, the location parameter of at least one component in the model should be fixed at the value of the anchor. For instance, a hypothesized anchor on $1/K$ should be tested with a model that has

$1/K$ fixed as the location parameter value for one of the mixture components, rather than allowing that parameter to be free for estimation. The most common example of this kind of model is the "zero-inflated" mixture model where 0 is the fixed location of a component distribution in a mixture model (see Lindquist & Gelman, 2009, for a recent example of such a model involving correlation coefficients).

Second, the researcher must decide whether to fix the precision parameter as well in the anchor-specific component distribution or estimate it. For maximum likelihood estimation (MLE), there are three options:

1. Assume infinite precision, that is, the anchor component distribution concentrates all of its mass at one point;
2. Assign a fixed precision parameter value for the anchor component distribution; or
3. Estimate the precision parameter for the anchor component distribution from the data.

The main drawback to the first option is that it rules out "near-misses" in the form of subjective estimates that are close to the anchor value. We recommend using it only when a strictly pointwise anchor is required by definition (e.g., as in a zero-inflated regression model) or when it is expected that normatively calibrated subjective estimates will be error-free. The second alternative can be set up to tolerate near-miss data, but it requires that the analyst impose an a priori precision parameter value. This option should therefore include a sensitivity analysis regarding the impact of the precision parameter value. The third option is viable if sufficiently stable estimates can be found. However, we have found in a number of applications that the low dispersion of sample data around an anchor renders precision parameter estimates unstable in these mixture models. That was the case for the examples presented in this paper, and so only the first and/or second options are employed in this paper.

A Bayesian approach admits another option, namely, an informative prior for the precision parameter instead of a fixed value. We illustrate this alternative in our examples, using Monte Carlo Markov Chain (MCMC) estimation. The appropriate sensitivity analysis here is similar to that for the MLE second option, that is, varying the input parameters of the informative prior and assessing the stability of the posterior estimates.

In the following sections, we provide three examples of the above models using probability judgment data.

4. Example 1: Pure Composition Shift

Smithson and Segale (2009) conducted an experiment on judged probabilities with a 2×2 factorial design. For the first experimental factor, half the participants were primed to think that there were two alternatives (the "case prime": Either Sunday will or will not be the hottest day of the week) and half were

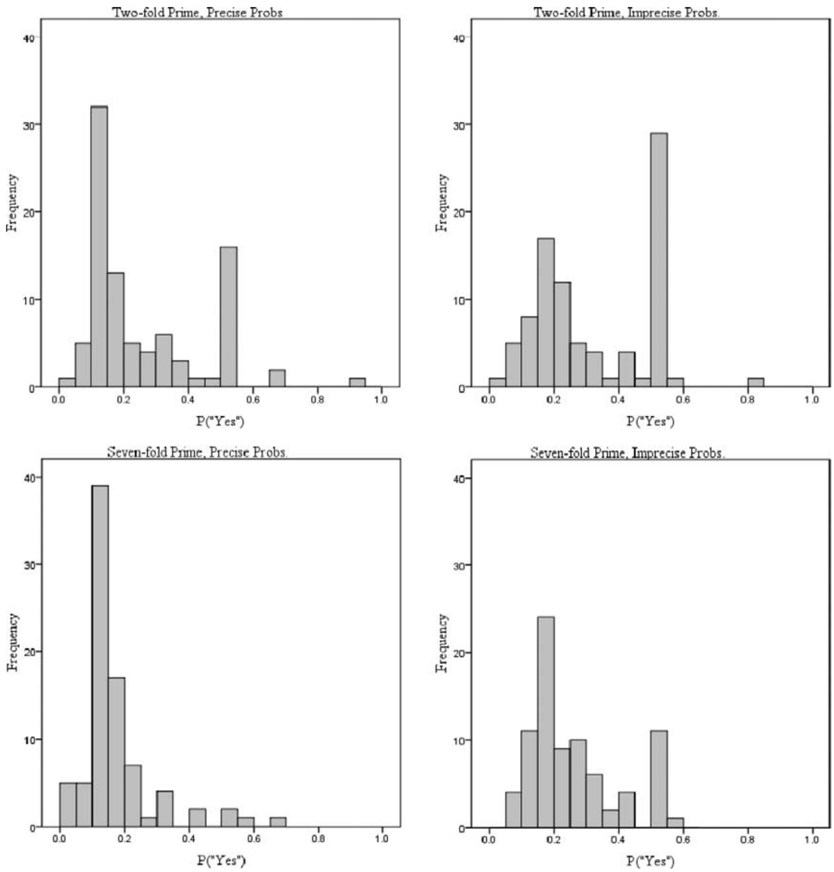

FIGURE 1. *Component distributions for class and case primes.*

primed to think there were seven alternatives (the "class prime": Of the 7 days in the week, Sunday will be the hottest). Their hypothesis was that participants' estimates of the probability that Sunday would be the hottest day would anchor around 1/2 in the first condition and 1/7 in the second. Participants were asked for probability estimates that the answer to this question would be "Yes," and for the probability that it would be "No." For the second experimental factor, half the participants were asked to give a precise probability as their estimate and the other half were asked to give a lower and upper probability as their estimate. The purpose of this manipulation was to determine whether people would be as prone to partition-priming effects when providing imprecise probabilities as when using precise probabilities. Imprecise probabilities were rendered comparable with precise probabilities by averaging the lower and upper probability estimates (see Smithson & Segale, 2009 for a justification of this approach over alternatives). Figure 1 shows the distributions of the "Yes" probabilities under the four combinations of conditions.

We present an analysis of the "Yes" probabilities that differs from the approach taken by Smithson and Segale. They used a mixture model in which location parameters in both component distributions were free, but the precision parameters were assumed to be equal. Here, we fix the anchor component distribution's mean at 1/2 and estimate the second component mean. Models in which both precision parameters were free and therefore separately estimated proved unstable. These models were unable to converge on an estimate of the precision parameter for the anchor component distribution. Therefore, we adopt the second option for handling the anchor component distribution's precision parameter, that is, varying its value systematically to determine the robustness of the model against different assignments thereof. Thus, we allow for "near misses," which are especially relevant to those in the imprecise probability estimation condition.

While the results do not differ substantively from the findings reported in Smithson and Segale, they do point to some pertinent considerations in this kind of modeling exercise. Table 1 displays the results for the best model obtained by Smithson and Segale (first row) and by our approach for the "Yes" probability judgments when the anchor component distribution's precision parameter δ_{20} is set to values of $-2, -4, -8$, and -12, that is, decreasing variance as the parameter decreases. For all such values, the best model was one with main effects on composition shift from partition (twofold vs. sevenfold) and elicitation method (precise vs. imprecise probability judgments). There were no effects of either experimental variable on location or precision, so this is an example of pure composition shift. Thus, the model may be written as

$$\begin{aligned}\log(\mu_{1i}/(1-\mu_{1i})) &= \beta_{10}\\ \mu_{2i} &= 1/2\\ \log(\phi_{1i}) &= -\delta_{10}\\ log(\phi_{2i}) &= -\delta_{20}\\ \log(\gamma_{1i}/(1-\gamma_{1i})) &= \theta_0 + \theta_1 Z_{1i} + \theta_2 Z_{2i},\end{aligned}$$

where $\delta_{20} = \{-2, -4, -8, -12\}$, $Z_{1i} = -1$ for the twofold partition and 1 for the sevenfold partition, and $Z_{2i} = -1$ for precise and $+1$ for imprecise probabilities. MLEs for this model were obtained using the NLMIXED procedure in SAS 9.2 (SAS Institute, 2009) and the CNLR procedure in PASW Statistics 18, with similar results. Standard errors of the estimates in SAS were obtained using the "truereg" option and in PASW via a bootstrap with 3,000 samples. Code and data for this and the other examples in this paper are avalable via a webpage at http://dl.dropbox.com/u/1857674/betareg/betareg.html. The Table 1 figures are from the PASW output.

TABLE 1

Results for Composition Shift Example

δ_{20}	$-$LL	$\chi^2(2)$	β_{10}	δ_{10}	θ_0	θ_1	θ_2
S.-S.							
−3.374	−256.032	33.114	−1.671 (0.033)	−3.374 (0.148)	−1.218 (0.151)	−0.566 (0.149)	0.465 (0.142)
MLE							
−2	−234.675	33.417	−1.687 (0.039)	−3.322 (0.194)	−1.019 (0.175)	−0.602 (0.164)	0.516 (0.149)
−4	−261.356	28.910	−1.614 (0.036)	−3.162 (0.139)	−1.418 (0.166)	−0.613 (0.158)	0.462 (0.147)
−8	−337.093	32.262	−1.506 (0.053)	−2.813 (0.181)	−1.846 (0.193)	−0.737 (0.184)	0.520 (0.163)
−12	−450.900	30.812	−1.502 (0.051)	−2.803 (0.172)	−1.844 (0.186)	−0.718 (0.176)	0.502 (0.161)

The $-$LL column contains the negative log likelihoods for the composition shift model and the $\chi^2(2)$ column displays the chi-square difference between this model and a null model without $\theta_1 Z_{1i} + \theta_2 Z_{2i}$. This column shows that there is no trend in the goodness-of-fit difference between the two models, suggesting that model improvement is stable under different values of δ_{20}. The remaining columns contain the parameter estimates, and these indicate converging values for each of them as δ_{20} decreases.

The primary differences between the Smithson–Segale analysis and ours pertain to the relative composition coefficients. By fixing the anchor component distribution mean at 1/2 and increasing its precision, a smaller proportion of the sample is included in it. For instance, the proportion of participants included in the anchor component distribution according to the Smithson–Segale results is $\exp(-1.218 + 0.566)/(1 + \exp(-1.218 + 0.566)) = .34$ for the twofold partition and .14 for the sevenfold, whereas when $\delta_{20} < -8$, these proportions have declined to .25 and .07, respectively. Inspection of Figure 1 lends plausibility to the notion that the 1/2 anchor distribution's mass is very nearly concentrated at 1/2 whereas the other component distribution is more dispersed, suggesting that δ_{20} could be smaller than the Smithson–Segale model suggests. Moreover, the negative log likelihoods in Table 1 indicate better likelihoods for the models with smaller δ_{20} values.

However, further investigation into goodness of fit does not yield reasons to prefer the alternatives to the Smithson–Segale model. Table 2 displays the observed means and variances for the four experimental conditions and their estimates from the Smithson–Segale model and models with $\delta_{20} = \{-2, -4, -8\}$. Starting with the means, none of the alternative models more closely fit the means than the Smithson–Segale model. Comparing observed with estimated means, the Smithson–Segale model yields a root mean squared error (RMS) = 0.015, while the $\delta_{20} = -2$ model RMS = 0.020, the $\delta_{20} = -4$ model RMS = 0.016, and the $\delta_{20} = -8$ model RMS = 0.022. Likewise, for the variances, the Smithson–Segale model RMS = 0.005, the $\delta_{20} = -2$ model RMS = 0.013, the $\delta_{20} = -4$ model RMS = 0.005, and the $\delta_{20} = -8$ model RMS = 0.010.

Turning now to the assignment of cases to component distributions, one way to evaluate the separability of component distributions is the ratio of each component density to the sum of the component densities at each data point. If these ratios are close to 0 or 1, then the component distributions are well separated. Here, we use $R_i = \text{Max}(f_{1i}(Y_i), f_{2i}(Y_i))/(f_{1i}(Y_i) + f_{2i}(Y_i))$ so that values near 1 indicate strong separation. The 10th percentile of R_i for the Smithson–Segale model is .957 and for the the $\delta_{20} = -8$ model it is .999, so there is relatively little difference among the models in component separation.

We now compare the "Yes" and "No" probability judgments to determine whether they differ in their mixture compositions. Here, we adopt a Bayesian MCMC approach that allows more flexibility in parameterization in one respect:

TABLE 2
Observed and Estimated Means and Variances

	Means			Variances		
	Twofold	Sevenfold		Twofold	Sevenfold	
Observed	0.250	0.168	0.210	0.032	0.014	0.025
Precise	0.306	0.245	0.277	0.031	0.018	0.025
Imprecise	0.278	0.206	0.243	0.032	0.017	0.026
$\delta_{20} = -2$	0.253	0.192	0.224	0.038	0.026	0.029
Precise	0.337	0.242	0.291	0.042	0.036	0.038
Imprecise	0.295	0.217	0.257	0.040	0.029	0.032
Smithson–Segale	0.243	0.191	0.218	0.027	0.015	0.018
Precise	0.313	0.230	0.273	0.035	0.025	0.028
Imprecise	0.278	0.210	0.245	0.031	0.018	0.022
$\delta_{20} = -4$	0.239	0.192	0.216	0.023	0.011	0.014
Precise	0.305	0.224	0.266	0.033	0.020	0.023
Imprecise	0.272	0.207	0.241	0.027	0.014	0.017
$\delta_{20} = -8$	0.234	0.195	0.215	0.016	0.005	0.007
Precise	0.295	0.217	0.258	0.028	0.012	0.015
Imprecise	0.264	0.206	0.236	0.020	0.007	0.009

We may use an informative prior instead of fixing the precision parameter, and we used $\delta_{20} \sim N(\mu_\delta, \sigma_\delta^2)$ with $\sigma_\delta^2 = 0.01$ and μ_δ taking several values in the interval $[-20, -8]$. The informative prior turned out to yield much the same effect as a fixed parameter does, so we do not discuss it any further. The models were estimated in WinBUGS 1.4.3 (Spiegelhalter, Thomas, Best, & Lunn, 2004) using a two-chain model, with a 5,000 iteration burn-in and estimates computed from the subsequent 10,000 iterations. Models with $\delta_{20} > -7$ failed to converge. The results reported here have δ_{20}, given a starting value of -8. The code, data, and initial values for this model are contained in the Appendix. This model captures the mean structure reasonably well (root mean squared error is RMS $= 0.014$ for the "Yes" judgments and RMS $= 0.075$ for the "No" judgments) and the variance structure also (RMS $= 0.005$ for the "Yes" judgments and RMS $= 0.016$ for the "No" judgments).

We focus on the composition part of the model. The composition coefficients may be written as

$$\log(\gamma_{1ki}/(1 - \gamma_{1ki})) = \theta_{0k} + \theta_{1k} Z_{1ki} + \theta_{2k} Z_{2ki},$$

TABLE 3
Relative Composition Parameters

Parameter	*M*	*SE*	2.5%	97.5%
ν_1	−1.895	0.184	−2.269	−1.556
ω_1	0.089	0.253	−0.409	0.597
ν_2	−0.719	0.177	−1.084	−0.385
ω_2	0.164	0.241	−0.300	0.652
ν_3	0.523	0.161	0.212	0.843
ω_3	−0.061	0.225	−0.497	0.378

for $k = 1, 2$ (1 = "Yes" and 2 = "No"); where

$$\theta_{jk} = \nu_j + (k-1)\omega_j, \text{for } j = 0, 1, 2.$$

The ω_j provides the tests of whether the compositions differ when $k = 1$ or 2. Thus, ω_0 tests whether the overall composition differs between the "Yes" and "No" probabilities, ω_1 tests whether the partition priming effect differs, and ω_2 tests whether the precise versus imprecise probability elicitation effect differs.

The composition parameter estimates and 95% credible intervals are shown in Table 3. All of the ω_j estimates are close to 0 and their credible intervals contain 0, so the evidence favors the claim that the mixture compositions of the "Yes" and "No" judgments are similar. Moreover, the ν_j estimates closely resemble their MLE counterparts for appropriate precision parameter runs. Finally, a sensitivity analysis (not presented here) demonstrates that the informative prior approach is quite stable under different assignments of prior means for the precision parameters, with the values of the other parameters varying less than they do under the previous fixed-value MLE approach. Thus, we have consistent findings pointing to a pure composition shift model for both the "Yes" and the "No" judgments.

Note that using extremity scores to model polarization in this example would mislead us into thinking that probability judgments are more polarized for the sevenfold prime, in the sense of scores being more extreme. Transforming the (0,1) scores into extremity scores by taking their absolute difference from 0.5 reveals that the mean probability for the sevenfold prime is significantly lower than that for the twofold prime. But of course, this effect is entirely due to the difference in the relative sizes of the component distributions, not a shift in the location of the distributions themselves. Earlier researchers on the topic of partition effects on probability judgments (e.g., Fox & Rottenstreich, 2003) interpreted similar findings to these in terms of mean differences, but Smithson and Segale's analysis and ours strongly suggest it is composition shift rather than location drift.

It is instructive to compare this example with another task from the same study. The Jakarta Stock Exchange (JSX) task (based on Fox & Clemen, 2005) has participants estimate the likelihood that the JSX will close on Friday in one of two ranges: “less than 500” versus “at least 500 but less than 1,000.” Participants were randomly assigned to a threefold prime condition (see below) or a sixfold prime condition:

Threefold Prime

“The Jakarta Stock Index (JSX) will close on Friday in one of these ranges:

1. less than 500,
2. at least 500 but less than 1000, or
3. at least 1,000.

(What is the probability that) the JSX will close in ranges (1) or (2)?”

Sixfold Prime

“The Jakarta Stock Index (JSX) will close on Friday in one of these ranges:

1. less than 500;
2. at least 500 but less than 1,000;
3. at least 1,000 but less than 2,000;
4. at least 2,000 but less than 4,000;
5. at least 4,000 but less than 8,000; or
6. at least 8,000.

(What is the probability that) the JSX will close in ranges (1) or (2)?”

In this task, there is no “correct” partition, so partitioning is arbitrary. As a result, Smithson and Segale found no evidence that a mixture distribution was superior to a single-distribution model. Instead, they found significant partition priming effects on the mean (a lower mean probability under the sixfold prime) and on precision (higher precision under the sixfold prime). The findings from these two tasks amount to a preliminary test of the proposition that when a correct partition is available, some people are unmoved by partition priming, but otherwise the influence of partitions results in a shift of the entire distribution.

5. Example 2: Testing a Location Drift Effect

From the Smithson and Segale data set, we combine data from two judgment tasks, each from two independent samples. The first is the Sunday Weather task in Example 1. The second task required participants to estimate the probability that Boeing’s stock would rise more than those in a list of 30 companies. For both tasks, half of the participants were asked to provide lower and upper probability estimates of how likely each event was to occur and how likely to not occur. As mentioned earlier, one of the normative requirements for coherency in lower–upper probability judgments is conjugacy in the sense that $\underline{P}(A) = 1 - \overline{P}(A^c)$, where $\underline{P}(A)$ is the lower probability of A and $\overline{P}(A^c)$ is the upper probability of the complement of A. A simple test of conjugacy therefore is $\underline{P}(A) + \overline{P}(A^c) = 1$, which provides a normatively specified anchor. For those respondents whose $\underline{P}(A) + \overline{P}(A^c) \neq 1$, we wish to investigate predictors of the behavior of this sum.

In this example, we model the location of $\left(\underline{P}(\text{No}) + \overline{P}(\text{Yes})\right)/2$ values (dividing by 2 to map the sum into the [0,1] interval) in this example as a function of the difference $\overline{P}(\text{Yes}) - \underline{P}(\text{Yes})$. Note that there is no necessary relationship between these two quantities (e.g., for conjugate lower and upper probability assignments). However, in the Smithson–Segale data, there are a large number of cases with simultaneously low values for $\underline{P}(\text{No})$ and $\overline{P}(\text{Yes})$ and cases with simultaneously high values for these two probability judgments. As $\overline{P}(\text{Yes}) - \underline{P}(\text{Yes})$ approaches 1, $\overline{P}(\text{Yes})$ becomes restricted to be near 1 and, conversely, when $\overline{P}(\text{Yes})$ is near 0 then $\overline{P}(\text{Yes}) - \underline{P}(\text{Yes})$ also must be close to 0. Therefore, for the participants whose probability judgments are not conjugate, their $\underline{P}(\text{No}) + \overline{P}(\text{Yes})$ values should be positively correlated with $\overline{P}(\text{Yes}) - \underline{P}(\text{Yes})$. If so, then deviations away from conjugacy would appear to be partly driven by imprecision in judged "Yes" probabilities. The scatterplot in Figure 2 suggests that this is true. The squares in this plot are the cases obeying the conjugacy rule, while the circles are the cases that did not. The circles display a clear positive relationship between the two variables.

Starting with an MLE model, there is a significant effect from task (Sunday Weather vs. Boeing Stock) on relative composition. Also, $\overline{P}(\text{Yes}) - \underline{P}(\text{Yes})$ does not predict composition and neither this difference nor task predicts precision. In the location submodel, we include main effects for both task and $\overline{P}(\text{Yes}) - \underline{P}(\text{Yes})$ and their interaction. The MLE model with sensitivity analysis may be written as follows:

$$
\begin{aligned}
\log(\mu_{1i}/(1-\mu_{1i})) &= \beta_{10} + \beta_{11}X_{1i} + \beta_{12}Z_{1i} + \beta_{13}Z_{1i}X_{1i} \\
\mu_{2i} &= 1/2 \\
\log(\phi_{1i}) &= -\delta_{10} \\
\log(\phi_{2i}) &= -\delta_{20} \\
\log(\gamma_i/(1-\gamma_i)) &= \theta_0 + \theta_1 Z_{1i},
\end{aligned}
$$

where $\delta_{20} = \{-4, -8, -10, -12, -14, -16, -18\}$, $X_{1i} = \overline{P}(\text{Yes}) - \underline{P}(\text{Yes})$, and $Z_{1i} = -1$ for the Boeing stock task and $+1$ for the Sunday task. We include the task effect on composition in a "null" model for comparison with the $\overline{P}(\text{Yes}) - \underline{P}(\text{Yes})$ and task effects in the location submodel. MLEs for this model

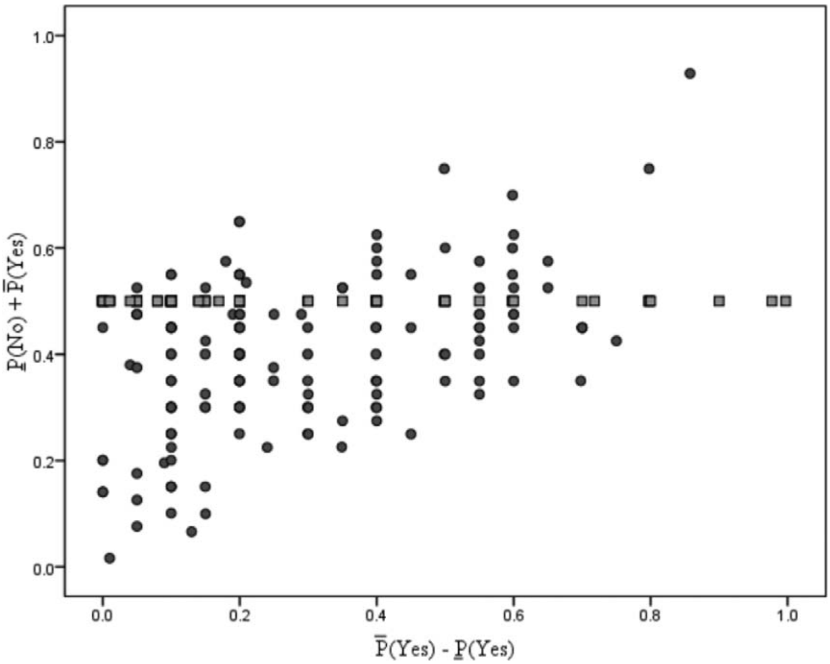

FIGURE 2. *Conjugacy scatterplot.*

were obtained using SAS 9.2 and PASW Statistics 18 and MCMC estimates were obtained via WinBUGS 1.4.3, with the same methods as in Example 1. The PASW results are reported here.

Unlike Example 1, model evaluation in this instance favors the models with greater precision (lower δ_{20}). The RMS error for the $\delta_{20} = -4$ model is 0.147 whereas for the $\delta_{20} = -8$ and $\delta_{20} = -12$ models it is 0.129. The observed mean is .433, $\delta_{20} = -4$ model estimated mean is .517 but the $\delta_{20} = -8$ and $\delta_{20} = -12$ model means are .474. The chief source of misfit is in the lower half of the distribution, and there the higher-precision models also perform better. For instance, the observed 25th percentile is .350 and the $\delta_{20} = -4$ model estimated 25th percentile is .475, whereas the $\delta_{20} = -8$ and $\delta_{20} = -12$ models yield .437 and .440, respectively. On the other hand, the observed median is .475 and the $\delta_{20} = -4$ model estimated median is .507 which exceeds the observed 75th percentile of .500, whereas the $\delta_{20} = -8$ and $\delta_{20} = -12$ models' medians are .467 and .468, respectively. Finally, the $\delta_{20} = -4$ model's component distributions are not only less well separated but also less appropriate than those of the more precise models. The $\delta_{20} = -4$ model's mean R_i is .86, whereas the $\delta_{20} = -8$ model's mean R_i is .95 and the $\delta_{20} = -12$ model's mean R_i is .998. Moreover, the $\delta_{20} = -4$ model's component distributions are not well separated for $Y_i > .4$

whereas the more precise models' distributions overlap substantially only in a narrow interval around .5, as should be the case.

Table 4 displays the parameter estimates for several appropriate MLE (PASW) and MCMC models (the latter used informative priors with $\delta_{20} \sim N(\mu_\delta, 0.01)$, where $\mu_\delta = \{-8, -14, -20\}$). In both kinds of model, there is a clear tendency for estimates to stablize as δ_{20} decreases. However, there is no trend in the $\chi^2(2)$ goodness-of-fit difference between the two MLE models, suggesting that model improvement is stable under different values of δ_{20}. An additional indication of model stability (not shown in Table 3) is the fact that in all three MCMC models the same cases were assigned to the posterior "normative" component distribution (the squares in Figure 2).

The location submodel β_{11} parameter for the main effect of $\overline{P}(\text{Yes}) - \underline{P}(\text{Yes})$ on the nonconjugate $\underline{P}(\text{No}) + \overline{P}(\text{Yes})$ values is positive, as expected. The task effect, β_{12}, is negative indicating that the Sunday Weather task nonconjugate $\underline{P}(\text{No}) + \overline{P}(\text{Yes})$ values are lower than those for the Boeing Stock task, as would be expected due to the finer partition in the Boeing Stock task. Finally, there is an interaction effect indicated by the β_{13} estimate, so that the positive relationship between the $\overline{P}(\text{Yes}) - \underline{P}(\text{Yes})$ and nonconjugate $\underline{P}(\text{No}) + \overline{P}(\text{Yes})$ values is stronger in the Sunday Weather task than in the Boeing Stock task. This relationship would be a mere artifact if $\overline{P}(\text{Yes})$ and $\underline{P}(\text{No})$ were independent of one another for nonconjugate judges, but in fact they are negatively related. Virtually nothing is known about what drives deviations from conjugacy of lower and upper probabilities, so this finding motivates further investigation to ascertain its psychological relevance and importance. Such investigations are, however, beyond the scope of this paper.

6. Example 3: A Three-Component Mixture Model

See, Fox, and Rottenstreich (2006) developed a partition-primed probability judgment task requiring participants to assign a probability to a transaction at a car dealership. Participants were informed that a car dealership sells two types of cars, coupes (two-door) and sedans (three-door), and employs four salespeople. Carlos deals exclusively in coupes while the remaining three (Jennifer, Damon, and Sebastian) deal in sedans. Participants were then told that a fictional customer wishes to trade in his current car for one of the same type. Participants were then asked one of two questions: "What is the probability that a customer trades in a coupe?" or "What is the probability that a customer buys a car from Carlos?"

The first question primes a twofold partition whereas the second primes a fourfold partition, so the hypothesis is that people will tend to anchor on 1/2 if asked about coupes and anchor on 1/4 if asked about Carlos. In an extension to See et al., Gurr (2009) included several individual differences measures of cognitive style preference. He combined two such scales, Need for Closure

TABLE 4

Results for Location Drift Example

δ_{20}	$-$LL	$\chi^2(2)$	β_{10}	β_{11}	β_{12}	β_{13}	δ_{10}	θ_0	θ_1
MLE									
−8	−287.717	49.265	−0.841	1.636	−0.181	0.746	−2.577	−0.822	0.479
			(0.129)	(0.314)	(0.134)	(0.326)	(0.128)	(0.213)	(0.212)
−10	−367.763	49.065	−0.833	1.620	−0.172	0.722	−2.564	−0.834	0.489
			(0.121)	(0.302)	(0.123)	(0.306)	(0.123)	(0.178)	(0.179)
−12	−454.387	49.410	−0.844	1.640	−0.169	0.718	−2.563	−0.797	0.469
			(0.122)	(0.301)	(0.125)	(0.309)	(0.124)	(0.175)	(0.169)
−14	−541.886	49.533	−0.847	1.648	−0.167	0.717	−2.562	−0.784	0.462
			(0.124)	(0.305)	(0.124)	(0.304)	(0.125)	(0.171)	(0.168)
−16	−629.702	49.579	−0.848	1.650	−0.167	0.717	−2.562	−0.779	0.460
			(0.125)	(0.305)	(0.125)	(0.307)	(0.124)	(0.171)	(0.168)
−18	−717.635	49.597	−0.849	1.651	−0.166	0.716	−2.562	−0.778	0.459
			(0.124)	(0.307)	(0.126)	(0.306)	(0.126)	(0.171)	(0.169)
MCMC									
−8			−0.834	1.622	−0.183	0.763	−2.539	−0.807	0.478
			(0.116)	(0.267)	(0.112)	(0.282)	(0.111)	(0.167)	(0.165)
−14			−0.848	1.653	−0.169	0.728	−2.530	−0.792	0.472
			(0.113)	(0.281)	(0.110)	(0.269)	(0.113)	(0.163)	(0.163)
−20			−0.853	1.659	−0.160	0.703	−2.529	−0.786	0.467
			(0.114)	(0.277)	(0.115)	(0.283)	(0.112)	(0.164)	(0.163)

Note: MCMC = Monte Carlo Markov chain; MLE = maximum likelihood estimation.

(Roets & Van Hiel, 2007) and Need for Certainty (Schuurmans-Stekhoven, 2005) because they were strongly correlated and appear to tap into much the same construct. Gurr investigated whether this combined scale (the NFCC) moderated the priming effect.

One hundred and fifty-five participants (108 females; 43 males; 4 unspecified) were recruited for the main study. These were undergraduate students at The Australian National University, some of whom obtained course credit in first-year Psychology for their participation in the study. Their ages ranged from 17 to 43 years ($M = 21.41$, $SD = 4.46$).

At first glance, this might seem to require a similar two-component mixture model to Example 1. However, participants were given information about both cars and salespeople, so it is plausible that some people might anchor on 1/2 and others on 1/4, regardless of the priming question. Therefore, our first comparison is between a two- and three-component mixture model. The two-component model assumes that each anchor is used exclusively in its respective priming condition, and these were modeled by uniform distributions of widths .002, .02, and .2. This model allows for a second component whose location and precision parameters are free. The model may be written as

$$\begin{aligned} f_{1i}(Y_i) &= \text{Uniform}(\mu_{1i} - d, \mu_{1i} + d) \\ \mu_{1i} &= 1/2 - Z_{1i}/4 \\ \log(\mu_{2i}/(1 - \mu_{2i})) &= \beta_{20} \\ \log(\phi_{2i}) &= -\delta_{20} \\ \log(\gamma_{1i}/(1 - \gamma_{1i})) &= \theta_{10} + \theta_{11} Z_{1i}, \end{aligned}$$

where d takes values $\{.001, .01, .1\}$ and $Z_{1i} = 0$ for the Car condition and 1 for the Salesperson condition.

The three-component model assumes that each anchor has its own component in both conditions and allows for a third component with free location and precision parameters:

$$\begin{aligned} f_{1i}(Y_i) &= \text{Uniform}(\mu_{1i} - d, \mu_{1i} + d) \\ f_{2i}(Y_i) &= \text{Uniform}(\mu_{2i} - d, \mu_{2i} + d) \\ \mu_{1i} &= 1/2 \\ \mu_{2i} &= 1/4 \\ \log(\mu_{3i}/(1 - \mu_{3i})) &= \beta_{30} \\ \log(\phi_{3i}) &= -\delta_{30} \\ \log(\gamma_{1i}/(1 - \gamma_{1i} - \gamma_{2i})) &= \theta_{10} + \theta_{11} Z_{1i} \\ \log(\gamma_{2i}/(1 - \gamma_{1i} - \gamma_{2i})) &= \theta_{20} + \theta_{21} Z_{1i}. \end{aligned}$$

MLEs were obtained for the models in this example in SAS 9.2 and PASW Statistics 18, using the same methods for standard error estimates as in Example 1.

TABLE 5
Parameter Estimates and Confidence Intervals for Example 3

			Confidence Interval	
Parameter	Estimates	*SE*	Lower	Upper
β_{30}	−0.299	0.176	−0.645	0.047
δ_{30}	−0.992	0.332	−1.643	−0.341
θ_{10}	0.239	0.232	−0.216	0.693
θ_{20}	−1.007	0.243	−1.21	−0.258
θ_{11}	−0.734	0.287	−1.57	−0.444
θ_{12}	0.484	0.2	0.092	0.875
θ_{13}	−0.681	0.266	−1.202	−0.161

The results reported here are for models with $d = .01$; other values of d produced similar findings. The log likelihood chi-square difference between these two models is large and significant ($\chi^2(2) = 124.350, p < .0001$), so the three-component model clearly is superior. Inspection of the three-component model parameter estimates reveals that the θ_{11} and θ_{21} estimates have similar magnitudes (−0.744 and 0.502, respectively). This suggests a restricted model in which $\theta_{11} = -\theta_{21}$, and it turns out that the fit for this model is almost identical to its unrestricted counterpart ($\chi^2(1) = 0.148, p = .700$). This result demonstrates that all of the effects pertain to a shift in relative composition between the 1/2 and 1/4 anchors. The effect of the prime on the third component is simply a by-product.

Now we incorporate the NFCC covariate into the new model, so that the composition submodel becomes

$$\log(\gamma_{1i}/(1-\gamma_{1i}-\gamma_{2i})) = \theta_{10} + \theta_{11}Z_{1i} + \theta_{12}Z_{2i} + \theta_{13}Z_{1i}Z_{2i},$$

$$\log(\gamma_{2i}/(1-\gamma_{1i}-\gamma_{2i})) = \theta_{20} - \theta_{11}Z_{1i} - \theta_{12}Z_{2i} - \theta_{13}Z_{1i}Z_{2i},$$

where Z_{2i} is NFCC transformed to a z-score variable.

This model significantly improves on the earlier one ($\chi^2(2) = 9.446$, $p = .009$). Its fit also is not significantly worse than an alternative model with separate parameters for the Z_{2i} and $Z_{1i}Z_{2i}$ terms in the second mixture component ($\chi^2(2) = 3.258, p = .196$). We therefore adopt it as the final model.

Table 5 shows the coefficients and bootstrap 95% confidence intervals from the PASW estimates for this model. As before, the negative θ_{11} coefficient indicates a shift from 1/2 to 1/4 as we move from the Car to the Salesperson condition. The positive NFCC coefficient, θ_{12}, indicates that in the Car condition there is a greater tendency for high-NFCC people to choose 1/2 but the positive θ_{13} coefficient tells us that this effect is eliminated in the Salesperson condition, presumably because so many people are choosing 1/4 in that condition. We may

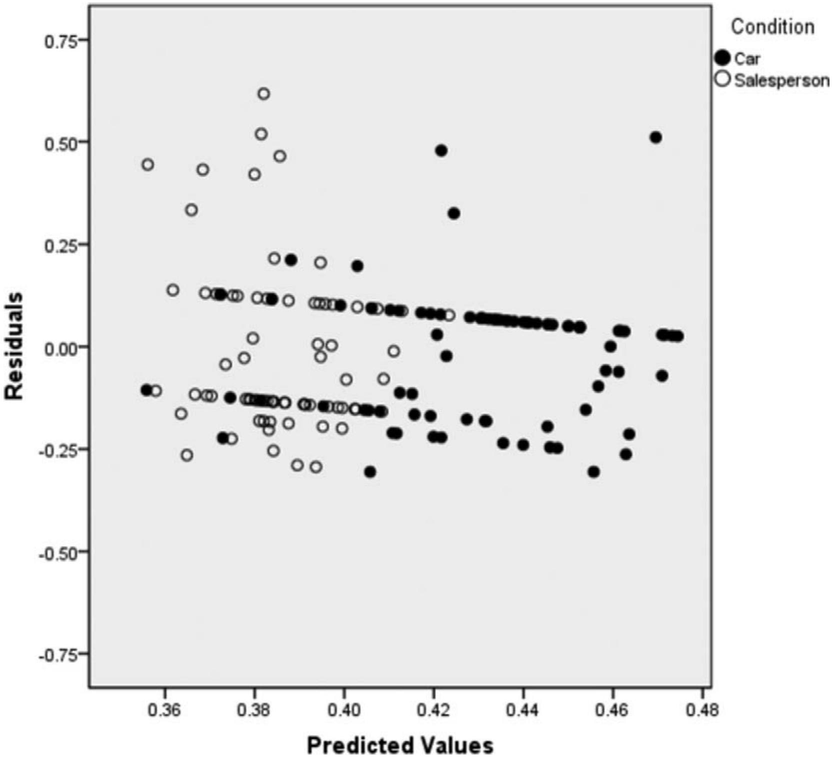

FIGURE 3. *Predicted values versus residuals.*

infer from this result that higher-NFCC people may be more susceptible to partition priming.

At the mean of NFCC, the model estimates of the proportions of respondents in the Car condition belonging in the 1/2-anchor, 1/4-anchor, and third component distributions are .480, .162, and .359, respectively, whereas the observed proportions are .481, .182, and .337. In the Salesperson condition, the model estimates are .259, .323, and .418 whereas the observed proportions are .260, .338, and .402. Thus, the model slightly underestimates the proportions for the 1/4-anchor components. However, it recovers the differences between proportions reasonably well. The observed composition shifts between the Car and Salesperson conditions are .481 – .260 = .221 and .182 – .338 = –.156, while the model estimates are .480 – .259 = .221 and .162 – .323 = –.161. Figure 3 displays the raw residuals plotted against predicted values. The model slightly underestimates the .5 and overestimates the .25 responses, but this is to be expected, given the component distributions. Seven outliers in the Salesperson condition and three in the Car condition skew the residuals somewhat. Nevertheless, both the root-mean and root-median squared residuals suggest that the model fits the Car condition (root-mean squared residual = .161, root-median squared residual = .092) somewhat better than the Salesperson condition (root-mean squared residual = .197, root-median squared residual = .133).

7. Conclusions

Although the examples of the mixture models discussed in this paper were confined to studies of judged probabilities, these models can be applied to a great variety of problems that currently lack appropriate methods. As observed in Verkuilen and Smithson (in press), doubly bounded psychological variables are quite common, including not only response scales such as sliders and visual analog scales but also derived measures such as the proportion of alloted time devoted to one task, or proportion of a credit card debt repaid in a given month. Mixture models may be extended to multilevel (mixed) mixture models (see Verkuilen & Smithson, in press, for a general characterization of multilevel models for beta-distributed dependent variables). Moreover, it is straightforward to extend these models to hierarchical modeling setups such as case-control comparisons on tests with binary items, via beta-binomial mixture models that account for overdispersion in both cases and controls.

These models can enhance the potential for theory testing and development in areas that concern the polarization or extremity of judgments or attitudes, priming, and anchoring effects. The potential benefits are threefold. First, the availability of appropriate distribution theory for handling the "problems" of skew, censoring, heteroscedasticity, and bimodality that characterize polarization and extremity enables these to be studied and modeled for the meaningful phenomena that they are. We note here that models that permit censoring (e.g., Tobit models) may be preferable when the bounds on the dependent variable's scale are arbitrary.

Second, the models presented here render theoretical terms more precise and operationally direct. For instance, the fact that relative composition and overlap (due to influences on first and/or second moments) can be modeled separately distinguishes among three kinds of polarization phenomena that heretofore have been ignored and/or conflated, and this should lead to more sophisticated theories of polarization, priming effects, and the like. Finally, the greater specificity in these models regarding types of anchors and polarization enhances the testability of theories about these phenomena by motivating or even requiring more specific (i.e., "bolder" in the Popperian sense) models and hypotheses. For example, the distinction between normatively specified anchors and nonnormative anchors provides a clue to and partial explanation for the existence of two distinct kinds of partition effects in probability judgments, as well as having clear implications for future research on this topic.

Appendix: Example 2 WinBUGS Code, Data and Initial Values

```
# The dependent variable is y
# x1 is the task (-1 for Boeing and 1 for Sunday)
# x2 is the difference between upper and lower P(Yes)
model
{
for(i in 1: N) {
y[i] ~ dbeta(omega[i], tau[i])
# We reparameterize the beta distribution
omega[i] <- mu[i]*phi[i]
tau[i] <- (1-mu[i])*phi[i]
# This is the location submodel
mu[i] <- exp(lambda[i])/(1+exp(lambda[i]))
lambda[i] <- (1-K[i])*(beta1 + beta2*x2[i] + beta3*x1[i] +
  beta4*x1[i]*x2[i])
# This is the dispersion submodel
phi[i] <- exp(-kappa[T[i]])
# This is the composition submodel
K[i] ~ dbern(P[i])
T[i] <- K[i] + 1
P[i] <- exp(m[i])/(1+exp(m[i]))
m[i] <- theta1 + theta2*x1[i]
}
beta1 ~ dnorm(0.0, 1.0E-6)
beta2 ~ dnorm(0.0, 1.0E-6)
beta3 ~ dnorm(0.0, 1.0E-6)
beta4 ~ dnorm(0.0, 1.0E-6)
theta1 ~ dnorm(0.0, 1.0E-6)
theta2 ~ dnorm(0.0, 1.0E-6)
kappa[2] ~ dnorm(-8.0,10.0)
kappa[1] ~ dnorm(0.0, 1.0E-6)
}
# Data
list(N = 242, y = c(0.5, 0.5498550725, 0.5, 0.5, 0.30057971, 0.5,
  0.200869565,
0.15101449249999999, 0.3504347825, 0.5, 0.5, 0.1011594205,
  0.45014492749999996,
0.45014492749999996, 0.400289855, 0.15101449249999999, 0.5,
  0.200869565,
0.4750724635, 0.5498550725, 0.5, 0.5, 0.6495652175, 0.5,
  0.2507246375,
0.400289855, 0.30057971, 0.5, 0.4501449275, 0.40028985499999997,
  0.5,
0.40028985499999997, 0.5, 0.3504347825, 0.30057970999999994,
  0.400289855, 0.5,
0.141043478, 0.5, 0.5, 0.30057971, 0.076231884, 0.475072464, 0.5,
0.40028985499999997, 0.1260869565, 0.5, 0.380347826, 0.5,
  0.400289855, 0.5, 0.5,
0.40028985499999997, 0.4501449275, 0.5, 0.3504347825,
```

```
  0.5249275365,0.599710145,
0.5,0.30057971,0.5,0.5,0.6495652175,0.151014493,0.400289855,
  0.5498550725,
0.3504347825,0.5,0.5,0.5,0.5,0.5,0.016405797,0.375362319,0.5,
  0.5,0.5,
0.30057971,0.5,0.5498550725,0.30057971,0.5,0.3504347825,0.5,
  0.5498550725,
0.2507246375,0.5,0.4501449275,0.5498550725,0.749275362,
  0.7492753624999999,
0.5,0.5,0.3504347825,0.5,0.5,0.30057971,0.4501449275,0.5,
  0.2507246375,
0.40028985499999997,0.5,0.2257971015,0.175942029,0.5747826085,
  0.5,0.5,
0.35043478250000004,0.475072464,0.69942029,0.5,0.195884058,
  0.400289855,
0.5,0.5348985505,0.375362319,0.5,0.5,0.2507246375,
  0.4501449275,
0.45014492749999996,0.3504347825,0.200869565,0.5,0.30057971,
  0.30057971,
0.40028985499999997,0.45014492749999996,0.5,0.5,0.5,0.5,0.5,
  0.4501449275,
0.35043478250000004,0.5,0.30057971,0.40028985499999997,0.5,
  0.5,0.30057971,
0.5,0.15101449249999999,0.400289855,0.5,0.5,0.5,0.3005797105,
  0.5,
0.45014492749999996,0.5,0.3504347825,0.30057971,0.5,
  0.3255072465,0.5,
0.0662608695,0.40028985499999997,0.5,0.45014492749999996,
  0.9287536235,
0.4501449275,0.30057971,0.4501449275,0.5,0.5,0.141043478,0.5,
  0.5,0.5,
0.2507246375,0.325,0.5,0.25,0.275,0.45,0.3,0.275,0.5,0.45,
  0.425,
0.325,0.25,0.5,0.225,0.55,0.5,0.35,0.45,0.425,0.45,0.45,
  0.475,0.5,
0.475,0.475,0.45,0.525,0.3,0.5,0.4,0.5,0.575,0.45,0.525,
  0.475,0.5,
0.525,0.5,0.5,0.5,0.525,0.35,0.35,0.475,0.475,0.55,0.45,0.4,
  0.475,
0.325,0.35,0.5,0.525,0.425,0.5,0.5,0.6,0.525,0.45,0.575,0.6,
  0.475,
0.475,0.575,0.625,0.525,0.225,0.5,0.5,0.1,0.625),
x1 = c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,-1,-1,-1,-1,-1,-1,-1,
-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,
  -1,
```

```
-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,
  -1,
-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,
  -1,
-1,-1,-1,-1),
x2 = c(0.049855072,0.19942029,0.0,0.0,0.19942029,
  0.19942028999999994,
0.0,0.099710145,0.099710145,0.099710145,0.0,0.099710145,
  0.099710145,
0.099710145,0.19942029,0.099710145,0.049855072,0.0,
  0.049855072,
0.099710145,0.39884057999999994,0.39884058,
  0.19942028999999994,0.79768116,
0.29913043500000003,0.29913043500000003,0.29913043500000003,
  0.0,0.19942029,
0.39884058,0.099710145,0.19942028999999994,0.5982608700000001,
  0.099710145,
0.09971014399999995,0.149565217,0.099710145,0.0,0.099710145,
  0.19942029,
0.19942029,0.049855072,0.049855072,0.139594203,
  0.19942028999999994,
0.049855073,0.498550725,0.039884058,0.498550725,0.19942029,
  0.0,
0.39884057999999994,0.19942028999999994,0.0,
  0.09971014399999995,0.19942029,
0.049855073,0.59826087,0.39884057999999994,0.099710145,
  0.19942028999999994,
0.19942028999999994,0.19942028999999994,0.149565218,
  0.19942029,
0.5982608700000001,0.39884057999999994,0.498550724,
  0.149565217,0.149565217,
0.19942029000000003,0.099710145,0.009971014,0.049855073,
  0.19942029,
0.97715942,0.099710145,0.099710145,0.79768116,
  0.19942028999999994,
0.19942029000000003,0.099710145,0.29913043500000003,
  0.19942028999999994,
0.09971014499999997,0.099710145,0.39884057999999994,
  0.39884058,
0.19942028999999994,0.498550724,0.79768116,0.099710145,
  0.099710145,
0.19942028999999994,0.19942029,0.19942029,0.29913043500000003,
0.19942029000000003,0.149565218,0.099710145,0.498550725,
  0.079768116,
0.34898550700000003,0.049855072,0.179478261,0.79768116,
  0.5982608700000001,
0.24927536200000003,0.049855073,0.59826087,0.19942029,
  0.08973913,
0.19942029000000003,0.149565218,0.20939130400000003,
  0.249275363,0.049855073,
0.009971015,0.19942029,0.099710145,0.099710145,
```

```
  0.6979710149999999,
0.099710145, 0.7179130439999999, 0.19942029000000003,
  0.099710145, 0.099710145,
0.39884057999999994, 0.29913043500000003, 0.19942029, 0.99710145,
  0.099710145,
0.19942029, 0.29913043500000003, 0.19942029, 0.079768116,
  0.19942029,
0.19942028999999994, 0.19942028999999994, 0.099710145,
  0.14956521700000003,
0.0, 0.099710145, 0.19942029, 0.099710145, 0.169507246, 0.0,
  0.14956521800000003,
0.079768116, 0.099710145, 0.099710145, 0.19942029, 0.099710145,
  0.009971015,
0.14956521800000003, 0.099710145, 0.129623189, 0.099710145,
  0.099710145,
0.099710145, 0.857507247, 0.099710145, 0.39884058,
  0.19942028999999994,
0.19942029, 0.049855073, 0.0, 0.099710145, 0.099710145,
  0.009971015, 0.099710145,
0.4, 0.4, 0.45, 0.4, 0.7, 0.3, 0.35, 0.35, 0.6, 0.55, 0.3, 0.3, 0.9,
  0.1, 0.4,
0.8, 0.4, 0.7, 0.75, 0.7, 0.45, 0.6, 0.5, 0.19, 0.55, 0.7, 0.55, 0.4,
  0.55, 0.5,
0.6, 0.65, 0.55, 0.55, 0.29, 0.5, 0.65, 0.1, 0.3, 0.7, 0.35, 0.5, 0.55,
  0.6,
0.55, 0.45, 0.2, 0.55, 0.25, 0.55, 0.6, 0.05, 0.6, 0.15, 0.4, 0.05,
  0.4, 0.35,
0.55, 0.55, 0.5, 0.2, 0.2, 0.4, 0.6, 0.15, 0.24, 0.04, 0.14, 0.15,
  0.4))
#Initial values for one chain
list(kappa = c(-2.0, -8.0), beta1 = -0.5, beta2 = 0.1, beta3 = -0.1,
  beta4 = 0.3, theta1 = 0.4, theta2 = 0.1, K = c(0, 1, 0, 1, 0, 1, 0, 1, 0,
  1, 0, 1, 0, 1, 1, 0,
1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0,
1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1,
0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0,
1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1,
0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0,
1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0,
1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1,
0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0,
1, 0, 1, 0, 1, 0, 1, 1, 0, 1))
```


Acknowledgments

The authors would like to thank David Budescu, Ken Mavor, an anonymous reviewer, and the editor for valuable discussions and suggestions.

References

Brauer, M., Judd, C. M., & Gliner, M. D. (1995). The effects of repeated expressions on attitude polarization during group discussions. *Journal of Personality and Social Psychology*, *68*, 1014–1029.

Downing, J. A., Judd, C. M., & Brauer, M. (1992). Effects of repeated expressions on attitude extremity. *Journal of Personality and Social Psychology*, *63*, 17–29.

Fieuws, S., Spiessens, B., & Draney, K. (2004). Mixture models. In P. de Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and non-linear approach* (pp. 317–340). New York, NY: Springer-Verlag.

Fox, C. R., & Clemen, R. T. (2005). Subjective probability assessment in decision analysis: Partition dependence and bias toward the ignorance prior. *Management Science*, *51*, 1417–1432.

Fox, C. R., & Rottenstreich, Y. (2003). Partition priming in judgment under uncertainty. *Psychological Science*, *14*, 195–200.

Gurr, M. (2009). *Partition Dependence: Investigating the Principle of Insufficient Reason, Uncertainty and Dispositional Predictors*. Unpublished Honours thesis: The Australian National University, Canberra, Australia.

Heinen, T. (1996). Latent class and discrete latent trait models: Similarities and differences. *Advanced quantitative techniques in the social sciences* (Vol. No. 9). Thousand Oaks, CA: Sage.

Lindquist, M. A., & Gelman, A. (2009). Correlations and multiple comparisons in functional imaging: A statistical perspective (Commentary on Vul et al., 2009). *Perspectives on Psychological Science*, *4*, 310–313.

Roets, A., & Van Hiel, A. (2007). Separating ability from need: Clarifying the dimensional structure of the need for closure scale. *Personality and Social Psychology Bulletin*, *33*, 266–280.

SAS 9.2. (2009). Cary, NC, U.S.A.: SAS Institute, Inc.

Schuurmans-Stekhoven, J. (2005). *The optimal ignorance model*. Unpublished PhD thesis. The Australian National University, Canberra, Australia.

See, K. E., Fox, C. R., & Rottenstreich, Y. S. (2006). Between ignorance and truth: Partition dependence and learning in judgment under uncertainty. *Journal of Experimental Psychology*, *32*, 1385–1402.

Smithson, M., & Segale, C. (2009). Partition priming in judgments of imprecise probabilities. *Journal of Statistical Theory and Practice*, *3*, 169–182.

Smithson, M., & Verkuilen, J. (2006). A better lemon-squeezer? Maximum likelihood regression with beta-distributed dependent variables. *Psychological Methods*, *11*, 54–71.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2004). *WinBUGS Version 1.4.3*. Cambridge, UK: Medical Research Council, Biostatistics Unit.

Verkuilen, J., & Smithson, M. (In press). Mixed and mixture regression models for continuous bounded responses using the beta distribution. *Journal of Educational and Behavioral Statistics*.

Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua advanced quantitative techniques in the social sciences* (Vol. 9). Thousand Oaks, CA: Sage.

Authors

MICHAEL SMITHSON is Professor in the Psychology Department at The Australian National University, Canberra A.C.T. 0200, Australia; Michael.Smithson@anu.edu.au. His research interests include judgment and decision making under uncertainty, nonlinear and generalized linear models, and fuzzy logic methods for the human sciences.

EDGAR C. MERKLE is an Assistant Professor in the Department of Psychological Sciences at the University of Missouri-Columbia, Columbia, Missouri, 65211, USA; ecmerkle@gmail.com. His work on this article was completed while he was at Wichita State University. His research interests include Bayesian methods, psychometric models, and subjective probability.

JAY VERKUILEN is Assistant Professor of Educational Psychology in the Graduate Center, City University of New York, 365 Fifth Avenue New York, NY 10016 U.S.A.; jverkuilen@gc.cuny.edu. His research interests include nonlinear and generalized linear models, item response theory and paired comparison models.

Manuscript received November 26, 2009
Revision received August 24, 2010
Accepted August 24, 2010