Never Say "Not:" Impact of Negative Wording in Probability Phrases on Imprecise Probability Judgments

Michael Smithson
The Australian National University
Michael.Smithson@anu.edu.au

David V. Budescu
Fordham University
budescu@fordham.edu

Stephen B. Broomell
Pennsylvania State University
broomell@gmail.com

Han-Hui Por
Fordham University
hanhui.p@gmail.com

Abstract

A reanalysis of Budescu et al.'s (2009) data on numerical interpretations of the Intergovernmental Panel on Climate Change (IPCC 2007) fourth report's verbal probability expressions (PE's) revealed that negative wording has deleterious effects on lay judgements. Budescu et al. asked participants to interpret PE's in IPCC report sentences, by asking them to provide lower, "best" and upper estimates of the probabilities that they thought the authors intended. There were four experimental conditions, determining whether participants were given any numerical guidelines for translating the PE's into numbers.

The first analysis presented here focuses on six sentences in Budescu et al. that used the PE "very likely" or "very unlikely". A mixed beta regression (Verkuilen & Smithson, in press) modelling the three numerical estimates revealed a less regressive mean and less dispersion for positive than for negative wording in all three estimates. Negative wording therefore resulted in more regressive estimates and less consensus regardless of experimental condition.

The second analysis focuses on two statements that were positive-negative duals. Appropriate pairs of responses were assessed for conjugacy and additivity. A large majority of respondents were appropriately super- and sub-additive in their lower and upper probability estimates. A mixed beta regression model of these three variables revealed that the $\underline{P}(A)$ and $\overline{P}(A^c)$ pairs adhered most closely to conjugacy. Also, the greatest dispersion occurred for $\underline{P}(A) + \overline{P}(A^c)$, followed by $P(A) + P(A^c)$. These results were driven by the dispersion in the estimates for the negatively-worded statement. This paper also describes the effects of the experimental conditions on conjugacy and dispersion.

Keywords. subjective probability, probability expression, elicitation, conjugacy, risk communication, climate change.

1 Introduction

The Intergovernmental Panel on Climate Change (IPCC) has provided reports that synthesize and assess information regarding scientific understanding of climate change phenomena and their potential impact. The fourth IPCC (2007) report utilizes verbal phrases to describe the uncertainties affiliated with its major claims. These phrases include positively- and negatively-worded probabilistic expressions (PE's, e.g., "very likely" and "very unlikely"). The guidelines for the IPCC fourth report provided its authors a numerical translation of the seven PE's they recommended for use in the report (Table 1). These guidelines also are included in the assessments and executive summaries.

Table 1: IPCC Probability Phrase Numerical Guides

Phrase	IPCC Range
Virtually certain	$> 99\%$
Extremely likely	$> 95\%$
Very likely	$> 90\%$
Likely	$> 66\%$
More likely than not	$> 50\%$
About as likely as not	$33\% - 66\%$
Unlikely	$< 33\%$
Very unlikely	$< 10\%$
Extremely unlikely	$< 5\%$
Exceptionally unlikely	$< 1\%$

Budescu, Broomell, and Por (2009) conducted an experimental study of lay interpretations of these PE's, using 13 relevant sentences from the IPCC report. Three sentences contained the PE "very likely," three others had "likely," three more had "more likely than not," three had "unlikely," and three used "very unlikely." PE's such as "very likely" are positively-worded PE's, whereas PE's such as "very unlikely" are negatively-worded PE's. Four examples are:

1. It is very likely that hot extremes, heat waves, and heavy precipitation events will continue to become more frequent.

2. Global average sea level in the last interglacial period (about 125,000 years ago) was likely 4 to 6 m higher than during the 20th century, mainly due to the retreat of polar ice.

3. Temperatures of the most extreme hot nights, cold nights and cold days are unlikely to have increased due to factors other than anthropogenic forcing.

4. It is very unlikely that hot extremes, heat waves, and and heavy precipitation events will not continue to become more frequent.

Budescu et al. asked 223 participants to interpret PE's in these sentences by providing lower, "best" and upper estimates of the probabilities that they thought the authors intended. Participants did so by using numerical sliders on a computer screen. Participants were randomly assigned to one of four conditions:

- Control: No numerical guide to the PE's
- Translation: Participants were shown the IPCC numerical translation guide to the PE's
- Wide: Each sentence contained its appropriate IPCC numerical translation guide
- Narrow: Each sentence contained a numerical translation that was a sub-interval of the IPCC translation range

Budescu et al. reported that participants' "best" estimates were more regressive (toward the middle of the unit interval) than the IPCC guidelines' stipulations, although less so in the Narrow and Wide conditions. The Narrow condition provided the largest improvement in the quality of responses over the Control condition.

Budescu et al. ensured that four of their target sentences included negatively-worded PE's, but they did not assess whether the valence of the PE's had any effects on participants' interpretations. Nevertheless, it is apparent from Figures 2-4 in their paper that the negatively-worded PE's yielded a greater spread of responses (i.e., less consensus) than the positively-worded phrases, and the median responses were more regressive. Both possibilities are worthwhile evaluating because of their implications for eliciting and communicating imprecise probability judgments. Indeed there is empirical evidence that "positive" and "negative" PEs induce different actions and interpretations (e.g. Teigen & Brun, 1999).

We model the lower ($\underline{P}(A)$), best ($P(A)$), and upper ($\overline{P}(A)$) probabilities simultaneously, via a mixed GLM for beta-distributed random variables (Smithson & Verkuilen, 2006; Verkuilen & Smithson, in press). A description of and rationale for this model are given in the Appendix, along with explanations of its parameters.

2 Positive Versus Negative Wording Effects

Responses to the three sentences using "very likely" and the three using "very unlikely" from Budescu et al. were modeled, with responses to the "very unlikely" statements subtracted from 1 to render them comparable to those from the "very likely" statements. Figure 1 shows boxplots of the resultant data. They indicate that there are differences in location and dispersion between the positive versus negative PE's, across the lower, best and upper estimates, and between experimental conditions.

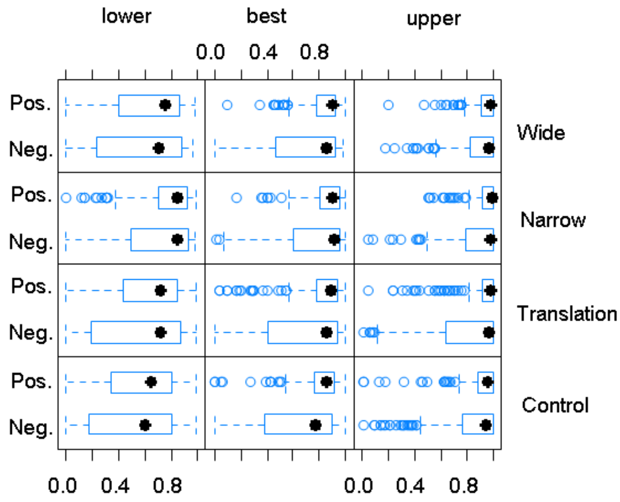

Figure 1: Boxplots of Estimates for Six Questions

We now describe the model of the effects shown in Table 2. The dependent vector consists of six sets of subvectors $\{y_{ij1}, y_{ij2}, y_{ij3}\} = \{\underline{P}(A)_{ij}, P(A)_{ij}, \overline{P}(A)_{ij}\}$, for $j = 1, \ldots, 6$. To respect the ordering $y_{ij1} \leq y_{ij2} \leq y_{ij3}$, we define $x_{i2} = 1$ for $y_{ijk} = y_{ij2}$ or $y_{ijk} = y_{ij3}$ and 0 otherwise, and $x_{i3} = 1$ for $y_{ijk} = y_{ij3}$ and 0 otherwise. We also restrict the regression coefficients for these dummy variables to be non-negative by exponentiating them. The "very likely" versus "very unlikely" predictor is $q_i = 1$ for "very likely" and 0 for "very unlikely". The experimental condition predictors are $t_{i1} = 1$ for the Translation condition, $t_{i2} = 1$ for the Narrow condition, $t_{i3} = 1$ for the Wide condition, and 0 otherwise. Using likelihood-ratio tests and AIC as guides, the best model is

$$\log\left(\frac{\mu_{ijk}}{1-\mu_{ijk}}\right) = \beta_0 + x_{2i}e^{\beta_1+\beta_2 q_i} + x_{3i}e^{\beta_3} + \beta_4 q_i \\ +\beta_5 t_{1i} + \beta_6 t_{2i} + \beta_7 t_{3i} + b_i, \tag{1}$$

where $b_i \sim N(0, e^{2u})$, and

$$\log(\phi_{ijk}) = \delta_0 + (\delta_1 + \delta_2 q_i)\, x_{2i} + (\delta_3 + \delta_4 q_i)\, x_{3i} + \\ (\delta_5 + \delta_6 t_{1i} + \delta_7 t_{2i} + \delta_8 t_{3i})\, q_i + \delta_9 t_{1i} + \delta_{10} t_{2i} + \delta_{11} t_{3i}. \tag{2}$$

The coefficients, standard deviations and confidence intervals are shown in Table 2.

Table 2: Mixed Model Parameter Estimates

			95% Confid.	Interval
Param.	Estim.	S.E.	Lower	Upper
		Location	Submodel	
β_0	-0.202	0.096	-0.391	-0.012
β_1	-0.354	0.081	-0.513	-0.196
β_2	0.472	0.089	0.297	0.647
β_3	-0.160	0.054	-0.266	-0.054
β_4	0.369	0.058	0.255	0.482
β_5	0.105	0.124	-0.139	0.349
β_6	0.768	0.139	0.494	1.042
β_7	0.343	0.134	0.078	0.607
u	-0.417	0.054	-0.524	-0.311
		Precision	Submodel	
δ_0	0.526	0.065	0.397	0.654
δ_1	0.319	0.066	0.189	0.448
δ_2	0.576	0.100	0.380	0.772
δ_3	-0.003	0.070	-0.141	0.135
δ_4	-0.272	0.095	-0.458	-0.085
δ_5	0.086	0.091	-0.093	0.265
δ_6	0.365	0.107	0.155	0.575
δ_7	0.707	0.125	0.460	0.953
δ_8	0.466	0.116	0.237	0.696
δ_9	-0.185	0.074	-0.332	-0.039
δ_{10}	0.459	0.087	0.288	0.629
δ_{11}	0.264	0.083	0.100	0.428

The location submodel's β_4 coefficient indicates that the positive statement probabilities were more extreme (less regressive) than their negative statement counterparts. This model's β_2 coefficient also shows that this effect is boosted for the "best" and upper estimates. Significant experimental condition effects occur only in the narrow and wide conditions. In both of those conditions responses are more extreme than in the control condition, and of course this effect is greatest for the narrow condition.

The precision submodel is somewhat more complex. The δ_1 coefficient indicates greater precision for the "best" probability estimates than for the lower probability estimates, and δ_2 suggests this is amplified for the positively-worded statements. However, the negative δ_4 coefficient suggests that this amplification does not hold for the upper estimates.

The positive-negative wording factor moderates the experimental conditions effects in the precision submodel. The interaction effect coefficients δ_7 and δ_8 amplify the greater precision effects from the narrow and wide conditions for the positively-worded sentences, while the δ_6 coefficient negates the lower precision in the translation condition for negatively-worded statements.

The model recovers the mean structure reasonably well. The observed and predicted means are shown in Table 3. The largest inaccuracies are a tendency to under-estimate the lower probability means, and the means for the negative PE's tend to have larger errors (RMS error = .045) than the positive PE's (RMS error = .029).

Table 3: Mixed Model Predicted and Observed Means

	control	treatment	narrow	wide
Negative:	"Very	Unlikely"		
Observed				
lower	.500	.552	.693	.580
best	.652	.686	.775	.702
upper	.825	.798	.863	.866
Predicted				
lower	.450	.476	.638	.535
best	.622	.647	.780	.699
upper	.794	.811	.893	.845
Error				
lower	-.051	-.076	-.055	-.048
best	-.028	-.039	.006	-.003
upper	-.031	.013	.028	-.021
Positive:	"Very	Likely"		
Observed				
lower	.562	.613	.769	.629
best	.809	.816	.856	.828
upper	.905	.912	.930	.927
Predicted				
lower	.542	.568	.718	.625
best	.784	.802	.887	.837
upper	.895	.905	.948	.923
Error				
Lower	-.021	-.045	-.051	-.004
best	-.024	-.015	.031	.009
upper	-.010	-.007	.019	-.003

3 Conjugacy

Two target sentences in Budescu et al. (2009) were positive-negative duals:

- Q1: It is very likely that hot extremes, heat waves, and heavy precipitation events will continue to become more frequent.

- Q12: It is very unlikely that hot extremes, heat waves, and heavy precipitation events will not continue to become more frequent.

This fact provides an opportunity to examine the relationships among subjective estimates of the lower and upper probabilities of A its complement A^c. Accordingly, this section assesses the responses to this pair of sentences for adherence to superadditivity for lower probabilities, subadditivity for upper proabilities, and the conjugacy rule for lower and upper probabilities.

The superadditivity requirement is $\underline{P}(A)+\underline{P}(A^c) \leq 1$, and the subadditivity requirement is $\overline{P}(A)+\overline{P}(A^c) \geq 1$. A large majority (83.4%) of the respondents' lower probabilities summed to less than 1, and an even larger majority (97.8%) of respondents' upper probabilities summed to more than 1.

Conjugacy is tested via the sums of appropriate pairs of responses, the criteria being
$\underline{P}(A) + \overline{P}(A^c) = 1$,
$P(A) + P(A^c) = 1$, and
$\overline{P}(A) + \underline{P}(A^c) = 1$,
where A^c denotes the complement of event A. Figure 2 shows the boxplots for the three sums and four experimental conditions. The medians all are quite close to 1 (conjugacy). However, there appear to be main effects on dispersion both for experimental conditions and the sums.

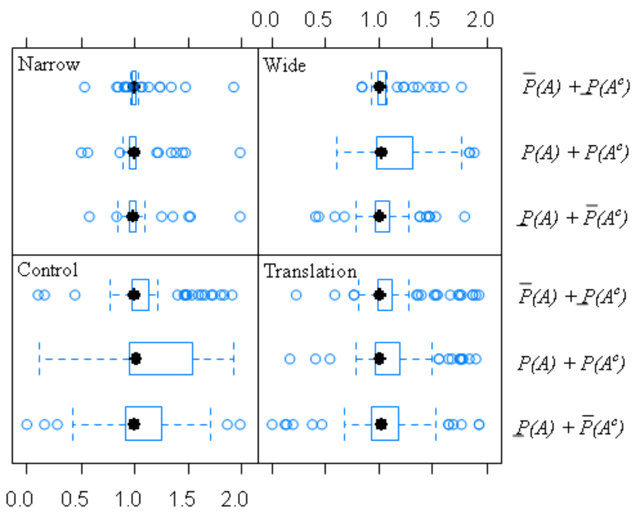

Figure 2: Boxplots of Sums

Turning to a model for the effects, for convenience the three sums described above were divded by 2, so that they lie in the unit interval. The dependent vector $\{y_{ij1}, y_{ij2}, y_{ij3}\}$ consists of the three sums in the order listed above, each divided by 2. We define $x_{i2} = 1$ for $y_{ijk} = y_{ij2}$ and 0 otherwise, and $x_{i3} = 1$ for $y_{ijk} = y_{ij3}$ and 0 otherwise. The experimental condition predictors are defined as before. In terms of likelihood-ratio tests and AIC the best model is

$$\log\left(\frac{\mu_{ijk}}{1-\mu_{ijk}}\right) = \beta_0 + \beta_1 x_{2i} + \beta_2 x_{3i} + b_i, \quad (3)$$

where $b_i \sim N(0, e^{2u})$, and

$$\log(\phi_{ijk}) = \delta_0 + \delta_1 x_{2i} + \delta_2 x_{3i} + \delta_3 t_{1i} + \delta_4 t_{2i} + \delta_5 t_{3i}. \quad (4)$$

The coefficients, standard deviations and confidence intervals are shown in Table 4.

Table 4: Conjugacy Model Parameter Estimates

			95% Confid.	Interval
Param.	Estim.	S.E.	Lower	Upper
		Location	Submodel	
β_0	0.140	0.054	0.035	0.246
β_1	0.126	0.036	0.054	0.197
β_2	0.060	0.031	-0.001	0.122
u	-0.388	0.056	-0.499	-0.278
		Precision	Submodel	
δ_0	2.401	0.170	-2.736	-2.066
δ_1	0.382	0.191	0.006	0.759
δ_2	1.148	0.259	0.638	1.657
δ_3	0.301	0.180	-0.053	0.656
δ_4	1.862	0.207	1.454	2.269
δ_5	0.622	0.189	0.249	0.996

The positive β_0 coefficient plus positive β_1 and β_2 show that the closest adherence to conjugacy in the means occurs for lower $\underline{P}(A)+\overline{P}(A^c)$. β_1 is largest so mean conjugacy is worst for $P(A)+P(A^c)$. The large positive δ_2 and moderate positive δ_1 coefficients show that the greatest precision occurs for $\overline{P}(A) + \underline{P}(A^c)$, followed by $P(A)+P(A^c)$. This result is being driven by the imprecision in the $P(A^c)$ estimates.

It turns out that there are no significant experimental condition effects in the location submodel but there are in the precision submodel. The positive δ_4 and δ_5 coefficients suggest that the narrow and wide conditions increase the precision of responses, the narrow condition substantially so.

This model also captures the mean structure well. The location submodel is slightly upward-biased, with the model estimates being about .02 higher than the observed values. However, this bias does not carry over into the differences between the means.

4 Discussion and Conclusions

In their summary and recommendations, Budescu et al. (2009) concluded that access to the IPCC numeri-

Table 5: Conjugacy Model Mean Structure

Conjugacy Sum	observed	predicted
$\underline{P}(A) + \overline{P}(A^c)$	1.052	1.070
$P(A) + P(A^c)$	1.120	1.132
$\overline{P}(A) + \underline{P}(A^c)$	1.076	1.100

cal translation table reduced individual differences in the interpretation of PE's to some degree. Our reanalysis reinforces this claim and their ensuing recommendation. Nevertheless, they also observed that the variability in respondents' estimates in all likelihood is greater than the actual amount of disagreement among the scientists whose views are encompassed by the relevant PE's. Budescu et al. based this assessment on their analysis of the "best" estimates. The reanalysis of the lower and upper probabilities in this paper suggests that the picture is even worse than their summary suggested.

They note, for instance, that 25% of the subjects interpreted "very likely" as having a "best" probability below 70%. The boxplots in Figure 1 show that in three of the four experimental conditions at least 25% of the subjects provided a lower probability of less than 50%. If we turn to "very unlikely" the picture is worse still. The Figure 1 boxplots indicate that in in three of the four experimental conditions about 25% of the subjects returned an upper probability for "very unlikely" greater than 80%!

Our reanalysis provides additional insights. Chief among these is the apparently deleterious impact of negatively-worded PE's on both the regressiveness of people's intuitive numerical translations of these PE's and on the consensus of such translations. Because beta GLMs are naturally heteroscedastic, it is both feasible to separate the effect of a shift in the mean from the effect of a shift in precision on variance. In this setting that separation has important implications regarding our assessment of the amount of variation across individuals in their intuitive numerical translations. More regressive estimates (i.e., further away from 0 or 1) results in greater variability, but that is an artifact of a shift in the mean response. Our results strongly suggest that negatively worded PE's also yield less precision, which results in greater variability that is not attributable to a mean shift.

Two other important findings have emerged regarding precision. First, it is worst for the lower (upper) probability estimates provided for "very likely" ("very unlikely"). But these are translations of the very thresholds identified in the IPCC numerical guides, as shown in Table 1. The effect also was greater for "very unlikely." Second, the narrow and wide conditions not only resulted in less regressive estimates (as Budescu et al. had originally concluded) but they also yielded greater precision, i.e., greater consensus beyond that due to less regressive estimates. This effect was greater for "very likely" than its negative counterpart.

The "pleasant surprise" in our analyses is the fairly strong adherence of subjective estimates to superadditivity, subadditivity, and the conjugacy rules. To our knowledge, only one other empirical assessment of adherence to conjugacy has been reported (Example 2 in Smithson, Merkle & Verkuilen, in press). In our sample, the medians in all conditions and for all three sums deviated no more than .1 from 1, i.e., conjugacy. A substantial majority of these sums were within .2 of 1 (from 52% to 86%). Moreover, both sums involving lower and upper probabilities were closer to conjugacy on average than $P(A) + P(A^c)$, which of course is just binary complementarity. This is striking because while many respondents would have been aware of the binary complementarity rule for classical probabilities, it is very unlikely that they would know about conjugacy. This may be a rather unusual instance where rational prescription coincides with human intuition. However, we urge caution in generalizing from these findings because they are based on only one pair of sentences. A systematic investigation into this matter is needed along the lines suggested below.

At least three avenues of future research are indicated by our findings here. First, the IPCC negatively-worded sentences contained a mixture of negatively-worded PE's and events (of the form "it is very unlikely that A will not occur"). Inspection of the data suggested that at least some respondents many have found these double-negatives especially confusing. Thus, the effect of negatively-worded PE's merits further investigation, most suitably via IPCC report sentences manipulated to incorporate positive and negative wording for various PE's and events crossed in a factorial design, as exemplified in Table 6. It is possible that the greater variability and more regressive means identified with the negatively-worded IPCC sentences are in good part due to double-negatives, but this cannot be determined via the study dealt with here.

Table 6: Factorial Design

Event	Probability phrase	
A	Likely that A	Unlikely that A
A^c	Likely that A^c	Unlikely that A^c

Second, alternative numerical guides could be compared with one another. The IPCC (2007) guides specified only one bound, leaving the other implicitly at either 0 or 1 as appropriate. For PE's conveying either very high or very low probabilities this seems natural, but for a middling PE such as "likely" an interval from .66 to 1 seems counter-intuitive not only for its width but also because it contains the prescribed interval for "very likely." The IPCC guidelines notwithstanding, it would be worthwhile to ascertain whether there is greater consensus in intuitive translations when the phrases refer to non-overlapping intervals instead of nested ones. Likewise, guides that include prescribed "best" probabilities could be compared with those containing only lower and upper values.

Finally, Budescu et al. suggested several influences on people's intuitive translations. For instance, those convinced about climate change tended to give higher estimates for PE's referring to climate change events or consequences. It is plausible that subjective probability judgments will be subject to confirmation bias, but this has yet to be investigated with respect to subjective imprecise probabilities.

5 Appendix

We begin by describing the mixed GLM employed in this paper. Let $y \in (0,1)$ be distributed Beta$(\mu\phi, (1-\mu)\phi)$, where $\mu = E(y)$ and ϕ is a precision parameter, such that $Var(y) = \mu(1-\mu)/(\phi+1)$ so $\phi = \frac{\mu(1-\mu)}{Var(y)} - 1$. As Smithson and Verkuilen (2006) argue, the Beta distribution is appropriate for modeling a random variable whose support is bounded at both ends, as in this case where the support is the unit interval. While it is not the only such distribution, it is very flexible and also has the attractive property of being parameterized in terms of a mean and a precision parameter. This characteristic renders the Beta distribution especially suitable for modeling the mean response (location) and dispersion simultaneously.

For a two-level model let $i = 1, \ldots, I$ index subjects and $j = 1, \ldots, J$ index observations within the ith subject, so there are $IJ = N$ total observations. A mixed beta GLM contains four matrices of regressors, $\mathbf{X}, \mathbf{Z}, \mathbf{V}, \mathbf{W}$. $\mathbf{X}$ and $\mathbf{V}$ are associated with the location and precision, respectively, so that $\mathbf{x}_i, \mathbf{v}_i$ are their ith row vectors of full rank (Typically they have a column vector $\mathbf{1}$ for an intercept). $\mathbf{Z}$ and $\mathbf{W}$ are the regressors for random effects $\mathbf{b}$ and $\mathbf{d}$, respectively. Then the location and precision submodels are

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_{ij}\mathbf{b}, \qquad (5)$$

$$\log(\phi_{ij}) = \mathbf{v}_{ij}\boldsymbol{\delta} + \mathbf{w}_{ij}\mathbf{d}. \qquad (6)$$

In this paper we restrict the random-effects models to random-intercept models for the location submodel with a normal mixing distribution.

Estimation was by maximum likelihood using the NLMIXED package in SAS 9.2. Maximum likelihood methods enable the use of both likelihood ratio tests for comparing models on the basis of goodness of fit, and Wald t- or z-tests for assessing the significance of individual coefficients in a model. The coefficients' standard errors used in the Wald tests may also be used in constructing confidence intervals for the coefficient estimates.

The location submodel coefficients in this model can be interpreted in a similar way to coefficients in a logistic regression, because the logit link typically is used in both. A positive (negative) β_j is the increase (decrease) in $\log(\mu_{ji}/(1-\mu_{ji}))$ per unit increase (decrease) in its covariate x_{ji}, so e^{β_j} can be interpreted as a multiplier of odds.

In the precision submodel, a positive (negative) δ_j coefficient is the increase (decrease) in $\log(\phi_{ji})$ per unit increase (decrease) in its covariate v_{ji}, so e^{δ_j} can be thought of as a multiplier of precision.

The variance of a Beta random variable is
$\sigma^2 = \mu_{ji}(1-\mu_{ji})/(\phi_{ji}+1)$,
so the variance is influenced both by the mean and precision parameters. This simply reflects the fact that as the mean approaches either 0 or 1, if the precision remains constant then the variance necessarily decreases. However, it is important to bear in mind that modeling precision is not equivalent to modeling the variance. Consequently, interpreting the effect of predictors on the variance may not be straightforward. A positive β_j, for instance, increases variance if it is shifting μ_{ji} from values below .5, but decreases variance if it is shifting μ_{ji} from values above .5.

Acknowledgements

The original survey work by Budescu, Broomell, and Por supported by the National Science Foundation under Grant No. 0345925.

References

[1] D. V. Budescu, S. Broomell, and H.-H. Por. Improving the communication of uncertainty in the reports of the Intergovernmental panel on climate change. *Psychological Science*, 20:299–308, 2009.

[2] Intergovernmental Panel on Climate Change. *Summary for policymakers: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate*

Change. Retrieved May 2010 from http://www.ipcc.ch/pdf/assessment-report/ar4/wg1/ar4-wg1-spm.pdf, 2007.

[3] M. Smithson and J. Verkuilen. A better lemon-squeezer? Maximum likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11:54-71, 2006.

[4] M. Smithson, Edgar C. Merkle and J. Verkuilen. Beta regression finite mixture models of polarization and priming. *Journal of Educational and Behavioral Statistics*, in press.

[5] K. H. Teigen and W. Brun. The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organizational Behavior and Human Decision Processes*, 80: 155-190, 1999.

[6] J. Verkuilen and M. Smithson. Mixed and mixture regression models for continuous bounded responses using the beta distribution. *Journal of Educational and Behavioral Statistics*, in press.