BRIEF REPORT

A Bayesian perspective on Likert scales and central tendency

Igor Douven¹

Published online: 27 July 2017
© Psychonomic Society, Inc. 2017

Abstract The central tendency bias is a robust finding in data from experiments using Likert scales to elicit responses. The present paper offers a Bayesian perspective on this bias, explaining it as a natural outcome of how participants provide point estimates of probability distributions over the items on a Likert scale. Two studies are reported that support this Bayesian explanation.

Keywords Bayesian modeling · Likert scales · Central tendency bias

The central tendency bias, which inclines participants to avoid the endpoints of a response scale and to prefer responses closer to the midpoint, is generally regarded as “one of the most obstinate” response biases in psychology (Stevens, 1971, p. 428). Various kinds of scales can give rise to this bias (see, e.g., Olkkonen, McCarthy, & Allred, 2014; Douven & Schupbach, 2015; Allred, Crawford, Duffy, & Smith, 2016), but central tendency is particularly well documented in data obtained using Likert-type questionnaires. The present paper is concerned with the central tendency bias as apparent in data from such questionnaires, and specifically with the question of why we find it there.

The explanation to be proposed is in line with the increasingly popular Bayesian approach to cognitive modeling

(Tenenbaum, Griffiths, & Kemp, 2006; Oaksford & Chater, 2007; Lee & Wagenmakers, 2013). According to the standard Bayesian picture, we come to assign so-called credences to all propositions expressible in our language at an early age, where formally these credences are supposed to be probabilities. Those probabilities are then to be “updated” as more and more evidence accrues, with updates being required to follow Bayes’ rule. Researchers from various quarters have defended this picture on normative grounds: this is how we ought to epistemically behave (see, e.g., Ramsey, 1926; de Finetti, 1970; van Fraassen, 1989; Jeffrey, 2004). Descriptively speaking, the picture can at best be an idealization (Earman, 1992). Nevertheless, while people are known to systematically deviate from Bayesian tenets in some contexts (e.g., Tversky & Kahneman, 1974; Baratgin & Politzer, 2007; Bes, Sloman, Lucas, & Raufaste, 2012; Douven & Schupbach, 2015), there is much recent research suggesting that the picture is descriptively not that far off. To the contrary, Bayesian approaches to human learning, optimal information search, argumentation, and conditional reasoning have predictively been very successful (Griffiths & Tenenbaum, 2006; Oaksford & Chater, 1994; Hahn & Oaksford, 2007; Over et al., 2007; Douven & Verbrugge, 2013; Douven, 2016). In the psychology of reasoning, these and other findings have led to the emergence of what has been dubbed “the New Paradigm” (Over, 2009; Elqayam & Evans, 2013), which has shifted attention from classical logic, which was traditionally seen as providing the norms of reasoning, to probability theory and, more broadly, Bayesianism.

A number of well-known biases have already been successfully explained along Bayesian lines (e.g., Griffiths et al., 2010), including a central tendency bias found in visual judgments (Huttenlocher, & Engebretson, 2000; Huttenlocher, Hedges, & Vevea, 2000; Crawford, Huttenlocher, & Hedges, 2006). According to Bayesian models of visual judgments,

All Supplementary Information as well as the data of the studies and the script for the analyses and simulations is available at https://osf.io/zkvuj/?view_only=d650fb95d36d46869161e38433dd758f.

✉ Igor Douven
igor.douven@paris-sorbonne.fr

¹ SND/CNRS/Paris-Sorbonne University, Paris, France

such judgments are influenced by prior probabilities concerning the stimuli, which tend to be highest for prototypical values. But this kind of model is unlikely to apply to Likert scale data, given that the midpoint of such a scale usually expresses some kind of neutrality, which is not the same as prototypicality. Indeed, expectations of typicality play no role in the Bayesian explanation to be put forward here. Rather, the first part of this explanation is that respondents to Likert-type questionnaires can be naturally thought to assign probabilities to the various response options, and that they will not always be fully certain about one of those options. Such uncertainty may pertain not only to factual matters beyond the participants' control but also to their own subjective attitudes (Abelson, 1988; Gross, Holtz, & Miller, 1995; Petrocelli, Tormala, & Rucker, 2007).

Likert-type questionnaires do not ask for probabilities; participants are to check exactly one of the options. From a Bayesian perspective, that is asking for a point estimate. An estimator of a random variable X based on another random variable Y is any function of Y that yields a value of X . In principle, that could be any function, but in practice, only a couple have received serious attention. The two best-known Bayesian estimators are the maximum a posteriori (or MAP) estimator and the least mean squares (or LMS) estimator. Given random variables X and Y and the observation that $Y = y$, the MAP estimate $\hat{x}_{\text{MAP}}$ of x is defined as either $\text{argmax}_x p_{X|Y}(x|y)$ or $\text{argmax}_x f_{X|Y}(x|y)$, depending on whether X is discrete or continuous; less formally, the MAP estimator returns the mode of the posterior distribution (the distribution of X conditioned on Y).¹ The LMS estimate $\hat{x}_{\text{LMS}}$ of x is defined as $\mathbb{E}[X|Y = y]$, which is either $\sum_x x p_{X|Y}(x|y)$ or $\int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx$, again depending on whether X is discrete or continuous; the LMS estimate equals the mean of the posterior distribution. The MAP estimate and the LMS estimate do not in general coincide; see the upper-left panel of Fig. 1 for an illustration.

Both estimators have attractive properties. Choosing the MAP estimate of X is known to minimize the error probability, that is, the probability that we are pronouncing a false value for X . On the other hand, choosing the LMS estimate is known to minimize the posterior mean squared error, in the sense that

$$\mathbb{E}[(X - \hat{x}_{\text{LMS}})^2 | Y = y] \leq \mathbb{E}[(X - \hat{x}_{\text{EST}})^2 | Y = y],$$

for any possible estimator EST of X based on Y .

It is often said that the MAP estimate is more appropriate when we are to choose among a number of competing hypotheses, whereas the LMS estimate is more appropriate when we are to estimate the value of a continuous parameter

(e.g., Bertsekas & Tsitsiklis, 2008, Ch. 8). At first, then, the MAP estimate might seem to apply in the case of selecting a response option on a Likert scale, a task which apparently confronts us with choosing among five or seven (or however many options there are) rival hypotheses.

On the other hand, Likert (1932) meant the scale type now named after him as a discretization of an underlying continuum, bounded by two endpoints. If that is also how participants view the Likert scale, then they may as well go by choosing the option that is closest to their LMS estimate. Let us say that participants who choose options in this way follow the rounding LMS rule, as opposed to the MAP rule, which has participants choose the option they deem most likely.

The second part of the proposed explanation is that respondents to Likert-type questionnaires tend to follow the rounding LMS rule rather than the MAP rule. For example, a participant whose probabilities for the options on a seven-point Likert scale are as represented in the upper-left panel of Fig. 1 would choose the midpoint—which is the option closest to her LMS estimate—and not the lower extreme, as would be dictated by the MAP rule.

To see how the foregoing might help to explain the central tendency bias, consider the upper-right panel of Fig. 1, which shows the outcomes of a more systematic comparison between responses that go by the MAP rule and responses that go by the rounding LMS rule. Specifically, it shows the choices of options under the rounding LMS rule, based on 10 million distributions randomly drawn from the uniform Dirichlet distribution (see Forbes et al., 2011, Ch. 13), in proportion to the choices of the same options under the MAP rule. The central tendency of the simulated results is impossible to miss. Only a small fraction of the distributions that under the MAP rule would have led to choosing one of the extreme options would have led to such an option under the rounding LMS rule. Conversely, more than 3.5 times the number of distributions that under the MAP rule would have led to a choice of the midpoint of the scale would have led to that choice under the rounding LMS rule.

It might be objected that, rather than providing some initial support for the Bayesian explanation of the central tendency bias proposed here, the simulation results undermine it. After all, while we do find that, in the relevant type of research, participants tend to avoid the extreme response options and show a preference for those more centrally located on the scale, we do not find this to the extreme extent suggested by the simulations.

Note, however, that it is not particularly realistic to assume that participants' probabilities for the various response options are typically random in the way probabilities sampled from a Dirichlet distribution are. The options on a Likert scale are *ordered*, with adjacent options also being conceptually closer than options further away on the

¹ It is to be noted that the MAP estimate need not be unique. However, here we take "MAP estimate" to be read as "MAP estimate, or MAP estimates if there is more than one"; below, we will be more precise on this point.

Fig. 1 *Upper-left panel* showing a probability distribution over Likert options (labeled 1 through 7), which gives rise to a difference between the MAP estimate (which is 1) and the LMS estimate (which is 3.67, and indicated by the *dashed line*). *Upper-right panel*: LMS choices of options, based on distributions randomly sampled from a

uniform Dirichlet distribution, in proportion to MAP choices, given the same distributions. *Bottom-left panel*: “realistic” probability distribution over Likert options. *Bottom-right panel*: LMS choices of options, based on random beta-binomial distributions, in proportion to MAP choices, given the same distributions

scale. For most participants, this order will be reflected in how likely they deem each of the options. Therefore, the probability distribution shown in the bottom-left panel of Fig. 1 appears more realistic as an example of a probability distribution over the options of a seven-point Likert scale than the one shown in the upper-left panel. There are multiple ways to model such more realistic probability distributions. One way is to use beta-binomial distributions, which are binomial distributions whose success parameter p itself follows a beta distribution, meaning that the parameters α and β which shape the beta distribution also shape the beta-binomial distribution (see Forbes et al., 2011, Sect. 8.7).

We used beta-binomial distributions instead of Dirichlet distributions for further simulations with the previously mentioned estimators, choosing α and β randomly and independently between 1 and 200 per run, and repeated the above simulations. This yielded the results shown in the bottom-right panel of Fig. 1, which exhibit a central tendency, but not of the extreme form present in the previous result.

These simulation results are only meant to bestow some initial plausibility on the proposed Bayesian explanation of the central tendency bias. For a more serious test of this explanation, we turned to experimental studies. The following presents the results of two studies that were designed to shed light on the question of whether people respond to surveys using Likert scales on the basis of (i) the probabilities they assign to the options on the scale, and (ii) the rounding LMS rule.

Both studies began by presenting participants with ten questions, each accompanied by a seven-point Likert scale. However, in contrast to standard such surveys, our studies explicitly asked the participants to assign probabilities to the various response options. This first part was then followed by a part whose only purpose was to distract the participants and to decrease the chance that, in the final part of the studies, the participants would remember what their responses to the first part had been. In both studies, the final part presented exactly the same questions that were asked in the first part, but this time the participants were asked to check one of the seven response options (in the first study) or to indicate their judgment by means of a slider (in the second study).

Comparing the data from the first parts of the studies with those of their final parts should provide insight into how participants' probabilities for the response options for a given question guided their summary judgment for that question. Our Bayesian hypothesis suggests that participants' responses in the final part of each study will be better predicted by what (on the basis of the first part) we can calculate to be their corresponding expectation, rather than by the option that in the first part received the highest probability.

Study 1: probabilities versus Likert scales

Method

Participants

There were 63 participants who were all recruited for a modest fee via CrowdFlower, where they were directed to the study on the Qualtrics platform. All participants were from Australia, Canada, the United Kingdom, or the United States. We excluded from analysis data from the 2.5% slowest and 2.5% fastest participants, then from non-native speakers of English, and participants who failed a validation question. (The validation question was taken from Aust, Diedenhofen, Ullrich, & Musch, 2013.) This left us with 55 participants (31 female; $M_{\text{age}} = 34$, $SD_{\text{age}} = 11$). These participants spent on average 681 s on the survey ($SD = 350$ s). Of these participants, 39 had a university education and 16 had only a high school or secondary school education.

Design and procedure

The first part consisted of ten questions concerning the press coverage of a number of events that were recent or ongoing at the time the survey was run (October 2016). Specifically, participants were asked how adequate they thought the press coverage was in their country, where they were given seven answer options, ranging from "Extremely adequate" to "Extremely inadequate," with all intermediate options labeled in the obvious way.² Participants were asked to indicate for each option, separately, exactly how strongly they believed it was the correct answer to the question, where their responses had to be on a scale from 0 to 100, and where they were told at the start, and reminded on each screen, that the numbers had to sum to 100 for each

question. (If the numbers did not add up to 100, the participant received a warning from the Qualtrics software.) Every question appeared on a separate screen in an order that was randomized per participant.

The second part consisted of two essay questions, which asked the participants to comment in five or six sentences on events that had been in the news (the events were not among those figuring in the first part of the study). As mentioned, this part was only meant to distract the participants before they began the final part.

The final part consisted of the identical questions that had been asked in the first part, except that now the answer scale was a real Likert scale, meaning that there were again seven answer options, labeled as in the first part of the study, but all that participants had to do was to check one of those options. Questions again appeared separately on screen in an order randomized per participant. See the Supplementary Information for the full materials.

Results

According to the first part of the Bayesian explanation, people attribute probabilities to the various options of a Likert scale. This claim would be challenged or at least trivialized, if we had found that our participants had consistently or nearly consistently attributed full probability (so, 100) to one of the response options. That turned out not to be the case: although 29% of the participants were certain of an option in all ten questions, 18% were not certain of any option in any question.³ On average, participants were certain of one of the options in 5.93 of the ten questions ($SD = 3.76$).⁴

The left panel of Fig. 2 summarizes the responses obtained in the final part of the survey. The events cited in the questions (e.g., the 2016 Olympics, Brexit, Hurricane Matthew, the ongoing war in Syria) all were being covered or had been covered rather widely by the main media outlets in the countries in which the survey ran. Therefore, it is not surprising to see a preponderance of positive responses. And in light of what is known about central tendency, it is not surprising, either, to see that the most positive option was still relatively sparingly chosen.

For the main part of the analysis, we conducted an ordinal logistic regression, using cumulative link models with a logit link function, as implemented in the ordinal package

²Labeling only the extremes is known to somewhat mitigate central tendency bias, but only by introducing another bias, which has participants favor labeled options because not having to figure out the meaning of a response option reduces processing effort; see Weijters, Cabooter, & Schillewaert (2010).

³To verify that our conclusions did not depend on the participants who were always certain of one option, we repeated the analysis to be reported below with those participants excluded. This led to qualitatively the same results; see the Supplementary Information.

⁴The R script that is available online contains a function, `prob.fnc`, that allows one to plot, for any given participant, that participant's probabilities assigned to the seven response options for each of the ten questions. The function works also for the responses from the second study.

Fig. 2 Histogram of the responses to the Likert scale questions from Study 1 (left; 1 codes the most positive response option, 7 the most negative one). Histogram of the responses to the slider questions from Study 2 (right)

(Christensen, 2015) for the statistical programming language R (R Core Group, 2016). The responses to the third part of the study were coded as 1 (for “extremely adequate”) through 7 (for “extremely inadequate”) and served as the dependent variable. From the first part of the study, we obtained participants’ rounded LMS estimates for each question as well as their MAP estimates. In those cases ($N = 60$) in which a participant’s MAP estimate was not unique, we picked the one closest to the choice in the corresponding question from the third part of the study. We thereby loaded the dice against our own explanation, which hypothesizes that participants’ responses to the questions in the third part are better predicted by their rounded LMS estimates from the first part than by the MAP estimates from that same part.

To see which provided the better predictions, we compared a model— CLM_{LMS} in the following—that had the rounded LMS estimates as independent variable with a model— CLM_{MAP} —that had the MAP estimates, determined as described, as independent variable. Likelihood ratio tests showed that, for both CLM_{LMS} and CLM_{MAP} , the optimal random effects structure included random intercepts and random slopes for participants but neither for questions. Each model was also compared with a constant-only model (the null model) that had the same random effects structure. Relevant model comparison results are displayed in Table 1.

The comparisons with the constant-only model indicated that rounded LMS estimates and MAP estimates are both reliable predictors of Likert choices. McFadden’s pseudo- R^2 values greater than .2 have been said to represent excellent model fit (McFadden, 1979), and we see that this threshold is met by both models. However, we also see that, according to this measure, CLM_{LMS} fits the data better than CLM_{MAP} . Furthermore, the AIC values reveal that rounded LMS estimates predict Likert choices much more reliably than MAP estimates do. Burnham and Anderson (2002,

p. 70) argue that a difference in AIC value greater than 10 means that the model with the higher value lacks any empirical support (AIC and BIC values quantify misfit, so lower is better). Table 1 shows that the difference in AIC value between CLM_{LMS} and CLM_{MAP} is more than 40 in favor of the former. Following a recommendation by Wagenmakers and Farrell (2004), we also calculated and compared Akaike weights, which showed that CLM_{LMS} is 7.76×10^9 times more likely to be the correct model than CLM_{MAP} (see also Vandekerckhove, Matzke, & Wagenmakers, 2015).

As a further test, we fitted a model— CLM_{FULL} —with both LMS estimates and MAP estimates as predictors, again with random intercepts and random slopes for participants. Likelihood ratio tests showed that the full model significantly improved on CLM_{MAP} — $\chi^2(35) = 69.40$, $p = .0005$ —but not on CLM_{LMS} , $\chi^2(35) = 23.86$, $p = .92$. As is seen in Table 1, CLM_{FULL} had a slightly better pseudo- R^2 value than CLM_{LMS} , but its AIC and BIC values were much worse, and even worse than those of CLM_{MAP} . Finally, inspection of the coefficients of CLM_{FULL} showed that, for all Likert response options, rounded LMS estimates were

Table 1 Comparison of cumulative link models

Model	LL	AIC	BIC	χ^2	R^2
CLM_{LMS}	−762.75	1605.50	1777.90	511.89	.25
CLM_{MAP}	−785.53	1651.10	1823.45	466.34	.23
CLM_{FULL}	−750.83	1651.65	1974.89	535.74	.26

Note. LL is the log-likelihood. AIC and BIC are the Akaike Information Criterion and Bayesian Information Criterion, respectively, which weigh model fit against model complexity, in related though somewhat different ways. The χ^2 values pertain to the likelihood ratio test comparing the given model with the null model; all three are significant at $\alpha = .0001$. The R^2 values reported are McFadden’s pseudo- R^2 values, which for any model M are defined as 1 minus (log-likelihood of M /log-likelihood of null model)

Fig. 3 Plots of models LM_{LMS} (left) and LM_{MAP} (right), with 95% CI bands. Data plotted with jitter, to enhance visibility

highly significant (all $ps < .0001$), whereas MAP estimates were significant for none of the response options (all $ps > .11$).⁵

Study 2: probabilities versus sliders

Another approach to testing our Bayesian explanation of the central tendency bias is to “undiscretize” the scale supposedly underlying the response options on a Likert scale and asking participants directly to pick a point on that scale that best reflects their judgment. That is what the second study did.

Participants

There were 64 participants in this study. Recruitment and testing proceeded the same way as in the first study. The exclusion criteria were also the same. In this study, applying those criteria left us with 54 participants (22 female; $M_{age} = 37$, $SD_{age} = 10$). These remaining participants spent on average 686 s on the survey ($SD = 382$ s). Thirty-five of them had a college education and 19 had only a high school or secondary education.

Design and procedure

The first and second parts of this study were identical to the first and second parts of Study 1. The third part again presented participants with the questions from the first part, but now

participants were asked to indicate their judgments about the adequacy of the news coverage of the various items by using a slider on a continuous scale (also known as a visual analogue scale; Reips & Funke, 2008), from -10 to 10 , where participants were reminded on each screen that -10 stood for “extremely inadequate” and 10 for “extremely adequate.” The starting position of the slider was always at the midpoint of the scale.

Results

Similar to what we found in the first part of the previous study, 26% of the participants were certain of an option in all ten questions, and 11% were not certain of any option in any of the questions.⁶ On average, participants were certain of one of the options in 5.61 of the ten questions ($SD = 3.58$). The plot in the right panel of Fig. 2 summarizes the responses to the final part of the current studies, the pattern being similar to that of the responses to the final part of Study 1. It is to be noted that while the sliders were on a continuous scale, the output from the Qualtrics software had finite (and rather limited) precision; hence the histogram instead of a density plot.

In the main part of the analysis, we used the *lme4* package for R (Bates, Mächler, Bolker, & Walker, 2015) to fit two mixed-effects linear models, both having participants’ responses to the slider questions as dependent variable, and one— LM_{LMS} —having the rounded LMS estimates based on the first part of study as independent variable and the other— LM_{MAP} —having the MAP estimates based on

⁵Daniel Heck suggested as an alternative analysis of the data, fitting mixed-effects linear models with LMS/MAP estimates considered as continuous variables. This analysis gave qualitatively the same results as the analysis reported here, the only exception being that the full model had a lower AIC value than the model with only LMS estimates as predictor. The R script in the Supplementary Information contains the code for the alternative analysis, for readers interested in further details.

⁶Here, too, the analysis to be reported was repeated with the “always certain” participants excluded. The reanalysis yielded qualitatively the same results as the analysis to be reported below; see the Supplementary Information.

Table 2 Comparison of mixed linear models

Model	LL	AIC	BIC	χ^2	R^2
LM _{LMS}	−1397.18	2803.05	2828.79	610.45	.80
LM _{MAP}	−1422.89	2854.52	2880.27	558.97	.78
LM _{FULL}	−1394.93	2809.85	2852.77	611.64	.81

Note. The χ^2 values pertain to the likelihood ratio test comparing the given model with the null model; they are significant at $\alpha = .0001$. The R^2 values reported are defined as the squared correlation between fitted values and responses from the slider questions

the same part of the study as independent variable (non-uniqueness was resolved in the same way as in Study 1). Both models included random intercepts as well as random slopes for participants but they included neither for questions, a choice justified on the basis of likelihood ratio tests. Figure 3 plots the models together with the responses to the slider questions versus the respective estimates. The models were compared with each other and with an intercept-only model with the same random effects structure. Table 2 gives the relevant outcomes.

Comparisons with the null model indicated that rounded LMS estimates and MAP estimates both reliably predict responses to the slider questions. AIC and BIC values again clearly favor the model with rounded LMS estimates as predictor. The R^2 values also show that LM_{LMS} better fits the data than LM_{MAP}, although the difference is not very large here. Calculating Akaike weights allows us to say that LM_{LMS} is 1.5×10^{11} times more likely to be the correct model than LM_{MAP}.

In addition to this, we compared LM_{LMS} and LM_{MAP} with a model—LM_{FULL}—that had both rounded LMS estimates and MAP estimates as predictors. Likelihood ratio tests showed that the full model improved significantly on LM_{MAP} but not on LM_{LMS}: $\chi^2(4) = 52.67$, $p < .0001$ and $\chi^2(4) = 1.19$, $p = .88$, respectively. This was buttressed by comparing AIC and BIC values; see Table 2. Finally, in LM_{FULL} rounded LMS estimates were a significant predictor ($p < .0001$) while MAP estimates were not ($p = .63$), contributing to the plausibility of the proposed Bayesian explanation of central tendency.

Conclusions

The central tendency bias is a robust finding in data from Likert-type questionnaires. What gives rise to this tendency is an open question. Are people influenced by an ill-grounded implicit belief that the truth tends to lie somewhere in the middle? Might they want to avoid appearing too extreme? Is the tendency due to a complex of processing problems (Krosnick, 1991)? To answer any of these

questions in the positive would seem to amount to portraying the central tendency bias as precisely that: a *bias*, something that, ideally, we would like to be absent from our data.

The present paper proposed a two-part Bayesian explanation of the central tendency bias, arguing that the bias is just what we should expect to see if people assign different probabilities to the response options on a Likert scale and then choose the option that is closest to their LMS estimate rather than the one that corresponds to their MAP estimate.

Results from two studies support this explanation. The first study showed that participants' rounded LMS estimates were better predictors of their Likert scale responses than their MAP estimates, and this conclusion held across a variety of criteria. The second study showed that when participants were given the freedom to pick a point on a continuum between the extremes of what in the first study was a seven-point Likert scale, their picks were reliably better predicted by their LMS estimates than by their MAP estimates.

Absent a normative argument to the effect that people *should* choose Likert scale options on the basis of the MAP rule rather than on the basis of the rounding LMS rule, our explanation of the central tendency bias actually implies that we have been wrong in thinking of it as a bias, threatening the validity of our research, and as something for which we should try to statistically correct (Baumgartner & Steenkamp, 2001). Supposing our explanation to be correct, we can conclude that data from Likert scale questions may be perfectly in order if they reveal a central tendency bias. We just have to know how to interpret them, namely, as reporting choices reflecting participants' rounded LMS estimates.

To be sure, while our best models showed good fit, there was room for improvement. That may indicate that we have uncovered only part of what underlies the central tendency bias in Likert scale data. One possibility is that there is a central tendency also in the probability distributions on which the point estimates are based. That central tendency could be due to the fact that, as Kahneman and Tversky (1979) have shown, people generally overestimate small probabilities and underestimate large probabilities. Whether correcting probabilities for this tendency and then basing point estimates on those corrected probabilities would lead to improved model fit is a question we leave for future research.⁷

Furthermore, we have looked at the best-known Bayesian estimators, the MAP estimator and the LMS estimator. But

⁷A problem one encounters here is that there is no unanimity on how best to model mathematically the probability distortions first observed by Kahneman and Tversky; for different proposals, see Tversky and Kahneman (1992); Camerer and Ho (1994); Wu and Gonzalez (1996); Prelec (1998) and Gonzalez and Wu (1999).

there is evidence suggesting that at least for some tasks people rely on different estimators, such as repeated sampling from their probability distribution and then taking the mean of the various samples (e.g., Acerbi, Vijayakumar, & Wolpert, 2014; Vul, Goodman, Griffiths, & Tenenbaum, 2014; Sanborn & Beierholm, 2016). Sanborn and Beierholm (2016) present data suggesting that which estimator people use may be task-dependent and may also differ from one person to another. To my knowledge, there are currently no studies on other Bayesian estimators directly concerned with responses from Likert-type questionnaires. We leave it as a further question for future research whether by taking into account other Bayesian estimators, and possibly also individual differences in point estimation, we can obtain more accurate predictions of such responses than are generated by the models presented in this paper.

Finally, our studies consisted of three parts, the middle part having the purpose of reducing, and ideally preventing, carry-over effects from the first part to the third, both of which present the same materials but ask for different types of responses. Internet-based surveys are limited in what can be done to distract participants: long distraction tasks can make an experiment prohibitively expensive or may lead to high attrition rates. It is therefore worth repeating the current studies or ones like them in a classroom setting in which what were now the first and third tasks can easily be spaced further apart in time, thereby giving greater assurance that carry-over effects have been avoided.⁸

References

- Abelson, R. P. (1988). Conviction. *American Psychologist*, 43, 267–275.
- Acerbi, L., Vijayakumar, S., & Wolpert, D. M. (2014). On the origins of suboptimality in human probabilistic inference. *PLoS Computational Biology*, 10, e1003661.
- Allred, S. R., Crawford, L. E., Duffy, S., & Smith, J. (2016). Working memory and spatial judgments: Cognitive load increases the central tendency bias. *Psychonomic Bulletin & Review*, 23, 1825–1831.
- Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, 45, 372–400.
- Baratgin, J., & Politzer, G. (2007). The psychology of dynamic probability judgment: Order effect, normative theories, and experimental methodology. *Mind & Society*, 6, 53–66.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Baumgartner, H., & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38, 143–156.
- Bertsekas, D. P., & Tsitsiklis, J. N. (2008). *Introduction to probability*, 2nd edn. Belmont: Athena Scientific.
- Bes, B., Sloman, S., Lucas, C. G., & Raufaste, E. (2012). Non-Bayesian inference: Causal structure trumps correlation. *Cognitive Science*, 36, 1178–1203.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multi-model inference: A practical information-theoretic approach*. Berlin: Springer.
- Camerer, C. F., & Ho, T. H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, 8, 167–196.
- Christensen, R. H. B. (2015). ordinal: Regression models for ordinal data. R package version 2015.6-28. <http://www.cran.r-project.org/package=ordinal/>.
- Crawford, L. E., Huttenlocher, J., & Engebretson, P. H. (2000). Category effects on estimates of stimuli: Perception or reconstruction? *Psychological Science*, 11, 280–284.
- Crawford, L. E., Huttenlocher, J., & Hedges, L. V. (2006). Within-category feature correlations and Bayesian adjustment strategies. *Psychonomic Bulletin & Review*, 13, 245–250.
- de Finetti, B. (1970). *Theory of probability* Vol. 2. New York: Wiley.
- Douven, I. (2016). *The epistemology of indicative conditionals*. Cambridge: Cambridge University Press.
- Douven, I., & Schupbach, J. (2015). The role of explanatory considerations in updating. *Cognition*, 142, 299–311.
- Douven, I., & Verbrugge, S. (2013). The probabilities of conditionals revisited. *Cognitive Science*, 37, 711–730.
- Earman, J. (1992). *Bayes or bust?* Cambridge: MIT Press.
- Elqayam, S., & Evans, J. St. B. T. (2013). Rationality in the new paradigm: Strict versus soft Bayesian approaches. *Thinking & Reasoning*, 19, 453–470.
- Forbes, C., Evans, M., Hastings, N., & Peacock, B. (2011). *Statistical distributions*. Hoboken: Wiley.
- Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, 38, 129–166.
- Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, 14, 357–364.
- Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, 17, 767–773.
- Gross, S., Holtz, R., & Miller, N. (1995). Attitude certainty. In Petty, R. E., & Krosnick, J. A. (Eds.) *Attitude strength: Antecedents and consequences* (pp. 215–245). Mahwah: Erlbaum.
- Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, 114, 704–732.
- Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, 129, 220–241.
- Jeffrey, R. (2004). *Subjective probability: The real thing*. Cambridge: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213–236.
- Lee, M., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling*. Cambridge: Cambridge University Press.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22, 5–55.
- McFadden, D. (1979). Quantitative methods for analyzing travel behaviour of individuals: Some recent developments. In Hensher, D., & Stopher, P. (Eds.) *Behavioural travel modelling* (pp. 279–318). London: Croom Helm.
- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101, 608–631.

⁸I am greatly indebted to Daniel Heck, Mike Oaksford, Henrik Singmann, Christopher von Bülow, and Eric-Jan Wagenmakers for valuable comments on previous versions of this paper.

- Oaksford, M., & Chater, N. (2007). *Bayesian rationality*. Oxford: Oxford University Press.
- Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. *Journal of Vision*, 14, 1–15.
- Over, D. E. (2009). New paradigm psychology of reasoning. *Thinking & Reasoning*, 15, 431–438.
- Over, D. E., Hadjichristidis, C., Evans, J. St. B. T., Handley, S. J., & Sloman, S. A. (2007). The probability of causal conditionals. *Cognitive Psychology*, 54, 62–97.
- Petrocelli, J. V., Tormala, Z. L., & Rucker, D. D. (2007). Unpacking attitude certainty: Attitude clarity and attitude correctness. *Journal of Personality and Social Psychology*, 92, 30–41.
- Prelec, D. (1998). The probability weighting function. *Econometrica*, 66, 497–527.
- Ramsey, F. P. (1926). *Truth and probability*. In *his Foundations of mathematics*, (pp. 156–198). London: Routledge.
- Reips, U.-D., & Funke, F. (2008). Interval-level measurement with visual analogue scales in Internet-based research: VAS generator. *Behavior Research Methods*, 40, 699–704.
- R Core Team (2016). R: A language and environment for statistical computing. Vienna. <http://www.R-project.org/>.
- Sanborn, A. N., & Beierholm, U. R. (2016). Fast and accurate learning when making discrete numerical estimates. *PLoS Computational Biology*, 12, e1004859.
- Stevens, S. S. (1971). Issues in psychophysical measurement. *Psychological Review*, 78, 426–450.
- Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, 10, 309–318.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representations of uncertainty. *Journal of Risk and Uncertainty*, 5, 297–323.
- Vandekerckhove, J., Matzke, D., & Wagenmakers, E.-J. (2015). Model comparison and the principle of parsimony. In Busemeyer, J., Townsend, J., Wang, Z. J., & Eidels, A. (Eds.) *Oxford handbook of computational and mathematical psychology*. Oxford: Oxford University Press.
- van Fraassen, B. C. (1989). *Laws and symmetry*. Oxford: Oxford University Press.
- Vul, E., Goodman, N. D., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, 38, 599–637.
- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11, 192–196.
- Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27, 236–247.
- Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science*, 42, 1676–1690.