Title
Measuring Belief Bias with Ternary Response Sets

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

Authors
Winiger, Samuel
Singmann, Henrik
Kellen, David

Publication Date
2018

Measuring Belief Bias with Ternary Response Sets

Samuel Winiger (s.winiger@psychologie.uzh.ch)
Henrik Singmann (singmann@psychologie.uzh.ch)
Department of Psychology, University of Zürich
Binzmühlestrasse 14/22, 8050 Zürich, Switzerland

David Kellen (davekellen@gmail.com)
Department of Psychology, Syracuse University
409 Huntington Hall, Syracuse, NY 13244, USA

Abstract

Belief bias in syllogistic reasoning refers to the finding that individuals are more likely to accept believable than unbelievable conclusions independent of their logical validity. Most theories argue that belief bias is driven by differences in reasoning processes between believable and unbelievable syllogisms. In contrast, Dube, Rotello, and Heit (2010) proposed that belief bias is solely an effect of response processes. We investigated belief bias without having to rely on response bias manipulations (Klauer, Musch, and Naumer, 2000) or confidence ratings (Dube et al., 2010). Instead, we added a third response ("I don't know") to the usual binary response set ("Yes"/"No"). This allowed us to test belief bias with a fully identified multinomial processing tree model, in a hierarchical Bayesian framework. We found evidence that the belief bias is driven by differences in response processes. Evidence for a difference in reasoning processes was inconclusive.

Keywords: belief bias; syllogisms; multinomial processing tree models

Introduction

Syllogisms are logical arguments that usually consist of two premises and a putative conclusion. In the *syllogism-evaluation task*, the task of the reasoner is to decide whether this conclusion follows necessarily from the premises (i.e., is valid) or not (i.e., is invalid). The validity of a conclusion is determined by the structure of the syllogism alone, independent of its actual content. Despite this property, reasoners tend to be influenced by prior beliefs about the conclusion. Evans, Barston, and Pollard (1983) crossed validity and believability (see Table 1 for an example) and found that believable conclusions are more likely to be endorsed than unbelievable ones. This effect was stronger for invalid syllogisms than for valid ones. Together, this effect of believability on endorsement rates for syllogisms constitutes the *belief bias* effect.

In a large set of studies, Klauer, Musch, and Naumer (2000) compared different accounts that try to explain the belief bias effect. What all accounts have in common, is that believability is assumed to affect *reasoning processes*: When confronted with a believable compared to an unbelievable syllogism, individuals are assumed to use the syllogistic structure and its content in different ways.

However, reasoning processes are not the only cognitive processes contributing to the performance in syllogism-evaluation tasks. The other main factor are *response processes*, which affect the general propensity to prefer one of the possible response options, independent of form and content of the syllogism. For example, a liberal reasoner is more likely to endorse syllogisms in general, whereas a conservative reasoner is more likely to reject them, compared to an unbiased reasoner.

Table 1: Example syllogisms.

	Believability	
Validity	Believable	Unbelievable
Valid	No oaks are jubs. Some trees are jubs.	No trees are punds. Some Oaks are punds.
	Therefore some trees are not oaks.	Therefore some oaks are not trees.
Invalid	No tree are brops. Some oaks are brops.	No oaks are foins. Some trees are foins.
	Therefore some trees are not oaks.	Therefore some oaks are not trees.

Klauer et al. (2000) showed that one cannot test different belief-bias accounts by comparing acceptance rates of conclusions. In order to establish diganostic tests, Klauer et al. developed an extended experimental and a formal measurement model – a multinomial processing tree model (MPT; Riefer & Batchelder, 1988) – to disentangle reasoning processes and response processes. Their model was a variant of the two-high threshold (2HT) model. It assumes that when reasoners are presented with a syllogism, there is a probability r with which they engage in reasoning and determine its true logical status (i.e., valid or invalid). If the reasoning step fails, with probability $1-r$, reasoners enter an uncertainty state in which they are forced to guess a response. Thus, this model assumes that reasoning can only lead to a correct response and errors are solely due to guessing. Further, response bias enters the model only in the guessing stage. Their model based analysis agreed with the notion put forward in essentially all existing accounts: Conclusion believability of a syllogism affected the reasoning processes.

Dube, Rotello, and Heit (2010) set out to answer the same questions using a different measurement model based on signal detection theory (SDT; Green & Swets, 1966). Their model assumes that the reasoning process leads to a continuous validity signal to which reasoners have access. When asked to evaluate the validity of a syllogism, reasoners compare its validity signal with an established response criterion that reflects their response bias. If the signal surpasses the criterion, reasoners endorse the syllogism, otherwise they reject it. Thus, this model can be seen as one instantiation of a probabilistic reasoning account (e.g., Oaksford & Chater, 2007). This stands in contrast with most established accounts of syllogistic reasoning, which assume that reasoning is performed on discrete entities such as sets, rules, or mental models (for an overview, see Khemlani & Johnson-Laird, 2012). Dube et al.'s model based analysis challenged all existing accounts of belief bias, as believability did not affect the reasoning processes. Instead, the belief bias could be explained by response processes only– a criterion shift indicating a response bias. Trippas Handley, and Verde (2013) replicated this finding for simple syllogisms (for which only one model of the premises needs to be constructed) but not for complex syllogisms (for which multiple models of the premises need to be constructed for a correct judgment of validity).

Despite the obvious differences between Klauer et al. (2000) and Dube et al. (2010) in terms of models and their assumptions, both studies had to address a technical problem complicating their modeling efforts. They wanted to estimate more model parameters than simple "Yes"/"No" datasets would allow for. More specifically, a single set of "Yes"/"No" responses to both valid and invalid syllogisms only provides two independent data points, thus only allowing the estimation of a maximum of two model parameters. However, as shown by Klauer et al. (2000), a comprehensive measurement model for syllogistic reasoning data requires the estimation of at least three parameters. A similar argument for the case of SDT was made by Dube et al. (2010, Experiment 1).

To increase the number of parameters that can be estimated, Klauer et al. (2000) used a response bias manipulation by collecting data from three groups of participants, each of which received different base rates of valid syllogisms. The low base rate group was instructed that only one sixth of the problems were valid. Medium, and high base rate groups were instructed that half and five sixth of the problems were valid, respectively. This procedure allowed Klauer et al. to circumvent the limitations of single "Yes"/"No" data sets. Specifically, they assumed that the base rate instructions would only affect response processes, but not reasoning processes. However, one potential limitation of this approach was that, by using three different groups, the model had to be applied to the aggregated data, thereby not allowing to account for individual differences.

Dube et al. (2010) used a different solution for making their full model identifiable. Instead of base rate instructions and equating parameters across groups, they expanded the response format from binary "Yes"/"No" responses to a 6-point confidence rating scale. They asked participants to provide a confidence judgment ranging from 1 to 3 (1 for low confidence, 2 for medium confidence, and 3 for high confidence) after each "Yes"/"No" response.

One problem with the SDT approach as introduced by Dube et al. (2010) is that it is unable to distinguish between a genuine response bias effect, a criterion shift, and a mathematically indistinguishable explanation via differences in reasoning processes, a distribution shift (Klauer & Kellen, 2011; Singmann & Kellen, 2014). Dube et al.'s conclusion that belief bias is merely a response bias effect hinges on the former interpretation of the results (for evidence of this interpretation, see Stephens, Dunn, & Hayes, 2017). This ambiguity regarding the reason for a specific results pattern is absent for the MPT model of Klauer et al. (2000). Unfortunately, formulating an MPT model for a 6-point confidence rating scale requires the specification of a response mapping function. This response mapping function can influence the conclusion, but is ultimately unrelated to the question of whether belief bias is driven by response processes or reasoning processes.

Here, we propose a minimal extension to the standard belief bias task that – similar to the approach by Dube et al. (2010) – does not require the use of a response bias manipulation. Neither is there a need for a 6-point confidence rating scale. We simply extend Klauer et al.'s (2000) model by one additional response category (Singmann, Klauer, & Kellen, 2013). Extending the response format from "Yes"/"No" to "Yes"/"I don't know"/"No" allows us to estimate all the necessary parameters.

One further difference between Klauer et al. (2000) and Dube et al. (2010) is the number of problems that each participants received. Klauer et al.'s participants solved eight to twelve problems. Dube et al. and following studies within an SDT framework (e.g., Trippas et al., 2013; Stephens et al., 2017) used considerably more problems per participant, usually 32 to 64. However, solving complex syllogisms requires considerable effort and concentration. Given such a large number of problems, it seems questionable whether participants invested an appropriate amount of effort for each syllogism. Therefore, we follow Klauer et al. and use only eight problems per participants.

One drawback of using a small number of problems per participant is that this precludes a model-based analysis on the individual level using a traditional maximum-likelihood approach. However, it is well-known that an analysis based on aggregated data for non-linear models such as discussed here can lead to severe aggregation artifacts (e.g., Klauer, 2010). Therefore, we analyze the individual-level data in a hierarchical-Bayesian framework using a partial-pooling approach (i.e., taking participants' similarities as well as their differences into account).

Extended Belief Bias MPT

The original 2HT model for syllogistic reasoning by Klauer et al. (2000) was tailored to binary "Yes"/"No" responses. The extended model has the three response categories "Yes","No", and "I don't know". Figure 1 shows the tree representation of the extended model, separately for believable and unbelievable syllogisms. In the following, we first describe the model in its general form which is equivalent for both believability conditions.

The extended belief bias MPT assumes a reasoning stage, followed by a response stage. With probability r a reasoner detects the correct logical status of a problem. In this case, the response stage simply consists of reporting this correct logical status of the problem, answering "Yes" for valid conclusions and "No" for invalid conclusions. With probability of $1 - r$ a reasoner does not detect the correct logical status of the problem. Please note that this reasoning parameter assumes reasoning processes to yield a correct answer, therefore reasoning processes that lead to errors (e.g., as proposed by reasoning accounts based on mental models) are not captured by this parameter.

If the correct logical status is not detected, an uncertainty state is reached and the response stage is entered. It is important to note that in this case the logical status of the problem is unknown and cannot influence the following decisions. With probability n a reasoner decides not to give a response about the logical status of the problem by selecting "I dont know". With probability $1-n$ a response is given despite having no information about the logical status of the problem. In this case, with probability g the response "Yes" is selected, and with probability $1-g$ the response "No" is selected. Parameters n and g capture response biases and, as mentioned above, incorrect reasoning processes, if the theory of interest proposes them. With these three parameters r, n, and g we build four processing trees, one for each type of problem:

- valid (v) syllogisms with believable (b) conclusions,
- invalid (i) syllogisms with believable (b) conclusions,
- valid (v) syllogisms with unbelievable (u) conclusions,
- invalid (i) syllogisms with unbelievable (u) conclusions.

For each problem type, we have a specific reasoning parameter r_{vb}, r_{ib}, r_{vu}, and r_{iu}. As argued above, the parameters representing response processes, n and g, cannot be influenced by the logical status of the problem, therefore we assume them to be equal for valid and invalid problems within believability conditions. They are however influenced by response biases. This leads to one parameter pair for each status of believability, n_b and g_b for believable problems and n_u and g_u for unbelievable problems. This of course assumes that other cues (e.g. premise believability, content specific effects) are controlled for across believability conditions.

The resulting four model trees are shown in Figure 1. A key assumption of the extension is that the n parameters are not a function of logical validity but only of response biases. Violations of this assumption would result in poor fit of the model to the data.

Statistical Analysis

Our analyses were performed in a hierarchical Bayesian framework. We used the latent-trait approach (Klauer, 2010) to fit the belief bias MPT as hierarchical Bayesian model in Stan (Carpenter et al., 2017).

Model fits were assessed via posterior predictive p-values, p_B, using the T_1 test statistic proposed by Klauer (2010). The T_1 statistic checks whether a model adequately describes the

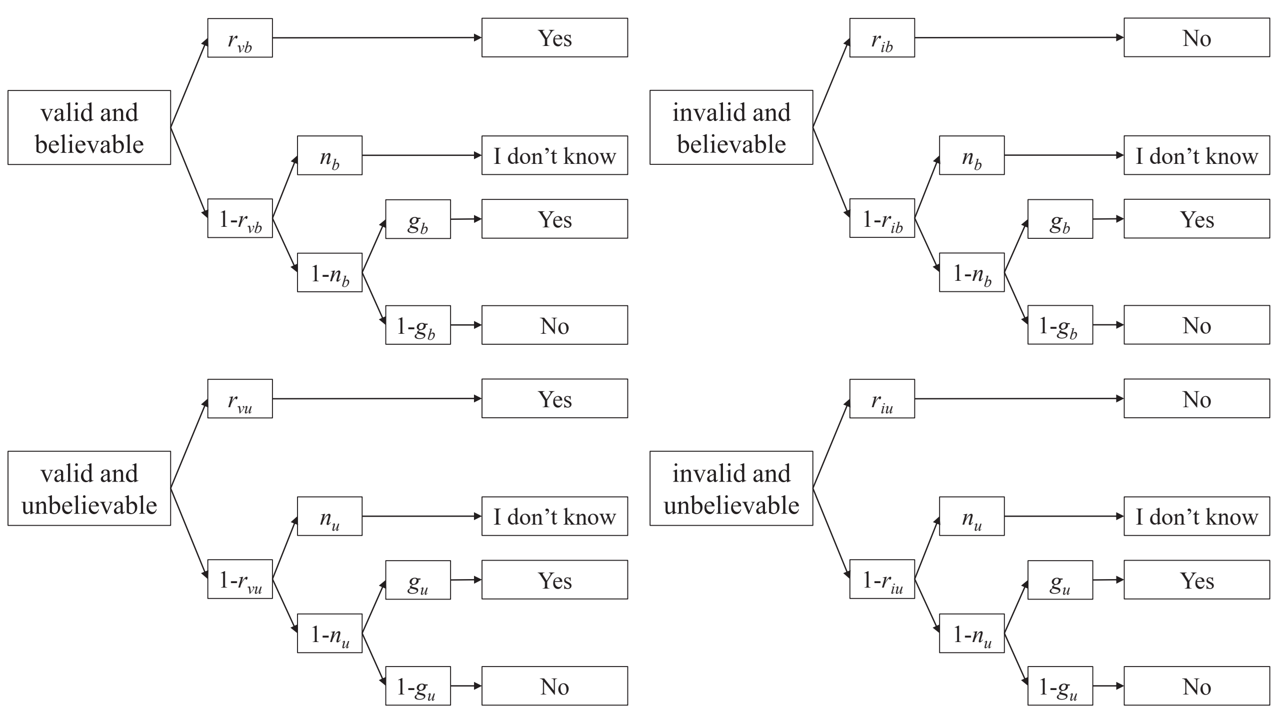

Figure 1: Four trees forming the extended belief bias MPT. r_{vb},r_{ib}, r_{vu}, and r_{iu} are the probabilities to reach the logically correct answer for the respective trees. n_b is the probability of answering "I don't know" for syllogism with believable conclusions (regardless of validity). n_u is the is the probability of answering "I don't know" for syllogisms with unbelievable conclusions. g_b is the probability of guessing that a syllogism with a believable conclusion is valid, thus responding with "Yes". g_u is the probability of guessing that a syllogism with a unbelievable conclusion is valid, thus responding with "Yes".

category frequencies, aggregated over participants. Small p_B values indicate poor model fits.

Our main research goal was to test differences in the model parameters between the believability conditions (e.g., between g_b and g_u). To this end, we reparameterized the model shown in Figure 1 to allow testing for differences. This was done by introducing a mean parameter $\bar{\theta}$ as well as a difference parameter δ_θ for each parameter $\theta \in \{r_v, r_i, n, g\}$. The condition specific parameter θ_b for the believable syllogisms and θ_u for unbelievable syllogisms were then obtained from those two parameters as

$$\theta_b = \bar{\theta} + \frac{1}{2}\delta_\theta \quad (1)$$

and

$$\theta_u = \bar{\theta} - \frac{1}{2}\delta_\theta. \quad (2)$$

This allowed us to test for differences between the two conditions by comparing prior and posterior distribution of δ_θ using the so-called Savage-Dickey density ratio (for a tutorial see Wagenmakers, Lodewyckx, Kuriyal & Grasman, 2010). The resulting Bayes factors quantify the evidence for (H_1) or against (H_0) a difference between believability conditions.

We followed Gronau, Wagenmakers, Heck, and Matzke (2017) and used a zero-centered normal prior for δ_θ. To account for the prior sensitivity of the Bayes factor, we explored different widths of the prior of δ_θ. This means, we fitted the model three times, each time with a different standard deviation σ_δ. σ_δ represents assumptions about the expected difference between the two believability conditions. We explored a narrow ($\sigma_{\delta\ narrow} \approx 0.52$), medium ($\sigma_{\delta\ medium} \approx 0.84$), and a wide ($\sigma_{\delta\ wide} \approx 1.28$) zero-centered normal prior. These priors represent the assumption of a small, medium, and large effects on the probability scale centered around 0.5 (see Gronau et al., 2017). More information about the model specification and parameter priors can be found in the model code on the Open Science Framework at: `https://osf.io/fsmvz/`

Experiment

The goal of the present experiment was to assess which cognitive processes are influenced by conclusion believability in a syllogism-evaluation task: reasoning processes, response processes, or both.

The Bayesian hierarchical framework of the analysis allowed us to collect only few data points per participant and capitalize on partial pooling. This seems especially valuable, given the high task difficulty of the syllogism-evaluation task with complex syllogisms.

Data, analysis scripts, and model codes can be found on the Open Science Framework at: `https://osf.io/fsmvz/`

Method

Design Logical validity (valid vs. invalid) and conclusion believability (believable vs. unbelievable) were manipulated within subjects.

Participants Four hundred thirty-seven participants were recruited for an online experiment via `crowdflower.com`. Participation was restricted to the following countries: UK, USA, Canada, Australia, and New Zealand. Participants were required to be native English speakers and at least 18 years old. From initially 437 participants, 83 were excluded resulting in 354 data sets. Exclusion criteria were as follows: Multiple IP entries (resulting in exclusion of all data sets from this IP; N = 30), missing IP entries (N = 31), and an average response time smaller than 2 seconds per syllogism (N = 29). Finally, we excluded participants who reported that they did not perform the task seriously (N = 10).[1] Participants could report this after the experiment and were informed that doing so would have no negative effect on their reward.

Each experimental session lasted less than 10 minutes. Participants were rewarded \$0.10 for participation and could win a performance based bonus. An additional \$0.50 was awarded if they responded correctly to at least 66% of the trials in which they selected the options "Yes" or "No". Trials in which they selected "I don't know" were neither counted as correct nor incorrect towards the bonus. This bonus system was aimed towards motivating participants to use the response option "I don't know".

Stimuli Twenty complex syllogistic structures were used. The structures were taken from Dube et al. (2010) experiments 1-3 and Klauer et al. (2000) experiments 3, 4, and 7. These syllogistic structures included the ones used by Trippas et al. (2013) and Stephens et al. (2017). The goal was to include all complex syllogistic structures that have recently been used in the literature on belief bias to minimize any effects that might be unique to a subset of these structures.

Participants were presented with four random syllogistic structures; two valid ones and two invalid ones. Each of these structures was presented once with a believable conclusion, and once with an unbelievable conclusion resulting in 8 unique syllogisms per participant.

Contents were chosen randomly from a list of forty rated pairs of category and exemplar (e.g., "trees" and "oaks"; Klauer et al., 2000; Dube et al., 2010; Ball, Phillips, Wade, & Quayle, 2006; Oakhill & Johnson-Laird, 1985; Quayle & Ball, 2000; Evans et al., 1983). Conclusion believability was manipulated by reversing the order of category and exemplar (e.g., "some trees are not oaks" and "some oaks are not trees"). Premises were linked with nonsense words to reduce effects of premise believability.

Procedure Participants were asked to decide whether a conclusion followed logically from the premises. Premises and conclusions were separated by a horizontal line. Responses were given by clicking on one of three buttons located right under the conclusions labeled "Yes", "I don't know", and "No". Selecting a response was immediately followed by the next problem.

Participants were instructed to assume that the premises

[1] Note that the exclusion criteria are not mutually exclusive

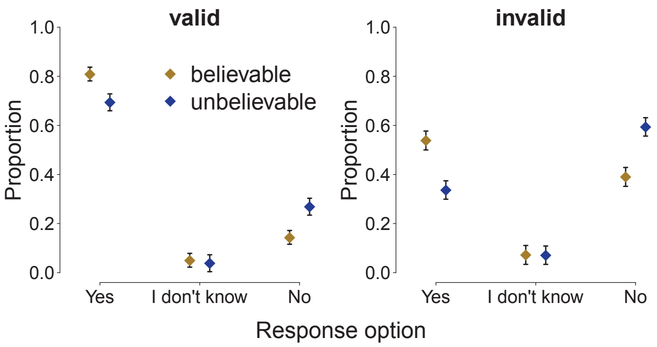

Figure 2: Response proportions to the three response options. Error bars represent the 95% multinomial confidence intervals (Sison & Glaz, 1995) based on data aggregated across participants.

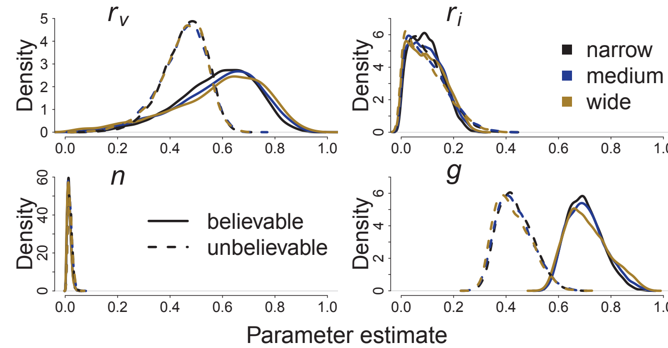

Figure 3: Posteriors of the group-level parameters r_v, r_i, n, and g. Different colors represent different prior widths for the difference parameters δ_θ.

were true and to endorse a conclusion only if it necessarily followed from the premises. We clarified that if a conclusion is possible, but not necessary, they were to select “No”. We also told them that they were to select “I don’t know”, if they could not decide whether a conclusion followed necessarily, but only if they seriously tried to evaluate the conclusion. We informed them that in the context of logic, the frequently used quantifier “some” means “at least one, possibly all”.

After the instructions, participants were shown an example problem, followed by a short version of the task instructions. Stimuli were divided into two blocks. This was not apparent to participants and there was no break between blocks. Each block consisted of four syllogisms using the four structures and random contents. Whether the conclusion was believable (or unbelievable) in the first or second block was randomly determined for each syllogistic structure anew. The order of syllogisms within each block was also randomized. The two stimuli blocks were followed by a performance feedback and a short demographic survey.

Bayesian Model We ran six MCMC chains with random start values, each running for 2000 samples and we discarded the first 1000 samples as warmup. Convergence of the MCMC chains was assessed via the $\hat{R}$ statistic ($\hat{R} < 1.04$). Results reported below are based on 6000 posterior samples.

Table 2: Bayes factors derived from the difference parameters

Parameter	r_v	r_i	n	g
narrow prior	$1.4^{\diamond}$	1.2	$2.1^{\diamond}$	232.0
medium prior	$1.5^{\diamond}$	$1.3^{\diamond}$	$3.1^{\diamond}$	70.3
wide prior	$2.0^{\diamond}$	$1.6^{\diamond}$	$4.5^{\diamond}$	16.1

Note. Bayes factors in favor of the null hypothesis (i.e., no difference between believability conditions) are marked with $^{\diamond}$ (i.e., $\text{BF}^{\diamond} = \frac{1}{\text{BF}}$).

Results

Response Frequencies Response proportions for each type of problem are shown in Figure 2. They exhibited a belief bias effect: The marginal proportion of responding with “Yes” was larger for believable conclusions than for unbelievable ones. The difference in responding with “Yes” between believability conditions occurred for both valid and invalid syllogisms. As expected, it was larger for invalid than for valid syllogisms. The proportions of responding with “I don’t know” seemed to be affected by neither validity nor believability.

Model Fit Overall model fit was good; for the model with the narrow prior $p_B = .51$, for the model with the medium prior $p_B = .50$, and for the model with the wide prior $p_B = .48$. This suggests that the model assumptions, such as n being unaffected by validity, were not grossly violated by the data.

Difference Analysis Posteriors of the group-level parameters are shown in Figure 3 and the Bayes factors derived from the difference parameters are shown in Table 2. We can see that the priors influenced the Bayes factors, but the results stayed qualitatively the same.

In contrast to Klauer et al. (2000), there was no evidence for an effect of believability on the reasoning parameters. More specifically, the evidence for a difference was inconclusive for both reasoning parameters r_v and r_i, $\frac{1}{3} < \text{BF} < 3$.

In terms of the response process parameters, the results tended to agree with Dube et al. (2010). For n, there was weak evidence for the absence of a difference between believability conditions, although only under the wide prior. The evidence under the medium and narrow prior was rather inconclusive. However, there was considerable evidence for a difference in response bias g over believability conditions, $\text{BF} > 16$. For the unbelievable conclusions participants appeared to be relatively unbiased with g around .4. For the believable conclusions participants showed a response bias towards the “Yes” response with g being clearly above .5 (at around .7).

Discussion

Comprehensive measurement models for syllogistic reasoning data require the estimation of at least three parameters, whereas binary "Yes/No" response formats provide only two independent data points (Klauer et al., 2000). We used a minimal extension to the binary syllogism-evaluation task by introducing a third response option, "I don't know" (Singmann et al., 2013). Although this third response option was rarely selected, it provided the means for estimating the parameters of the extended belief bias MPT. Similar extensions can be useful for investigating cognitive processes in other domains (e.g., source memory; Kellen, Singmann, & Klauer, 2014).

Here, we found converging evidence for the results of Dube et al. (2010) and the interpretation of their model-based analysis; the belief bias effect can be explained by a change in response bias (see also Stephens et al., 2017; Trippas et al., in press). However, we cannot conclude that belief bias is solely a result of response processes, as the evidence for or against differences in reasoning processes was inconclusive. Then again, considering our somewhat large number of participants we can assume that if there was a difference in reasoning processes it would probably be rather small.

References

Ball, L. J., Phillips, P., Wade, C. N., & Quayle, J. D. (2006). Effects of Belief and Logic on Syllogistic Reasoning. *Experimental Psychology*, *53*(1), 77–86.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1).

Dube, C., Rotello, C. M., & Heit, E. (2010). Assessing the belief bias effect with ROCs: It's a response bias effect. *Psychological Review*, *117*(3), 831–863.

Evans, J. S. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*(3), 295–306.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Gronau, Q. F., Wagenmakers, E.-J., Heck, D. W., & Matzke, D. (2017). A simple method for comparing complex models: Bayesian model comparison for hierarchical multinomial processing tree models using Warp-III Bridge sampling. *PsyArXiv: 10.17605/OSF.IO/YXHFM*.

Johnson-Laird, P. N., & Byrne, R. (1991). *Deduction* (Vol. xii). Hillsdale, NJ, US: Lawrence Erlbaum.

Kellen, D., Singmann, H., & Klauer, K. C. (2014). Modeling source-memory overdistribution. *Journal of Memory and Language*, *76*, 216–236.

Khemlani, S., & Johnson-Laird, P. N. (2012). Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, *138*(3), 427–457.

Klauer, K. C. (2010). Hierarchical Multinomial Processing Tree Models: A Latent-Trait Approach. *Psychometrika*, *75*(1), 70–98.

Klauer, K. C., & Kellen, D. (2011). Assessing the belief bias effect with ROCs: Reply to Dube, Rotello, and Heit (2010). *Psychological Review*, *118*(1), 164–173.

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, *107*(4), 852–884.

Newstead, S. E., Pollard, P., Evans, J. S. B. T., & Allen, J. L. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition*, *45*(3), 257–284.

Oakhill, J., Johnson-Laird, P. N., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition*, *31*(2), 117–140.

Oakhill, J. V., & Johnson-Laird, P. N. (1985). The effects of belief on the spontaneous production of syllogistic conclusions. *The Quarterly Journal of Experimental Psychology Section A*, *37*(4), 553–569.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: the probabilistic approach to human reasoning*. Oxford UP.

Quayle, J. D., & Ball, L. J. (2000). Working memory, metacognitive uncertainty, and belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology: Section A*, *53*(4), 1202–1223.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*(3), 318–339.

Singmann, H., & Kellen, D. (2014). Concerns with the SDT approach to causal conditional reasoning: a comment on Trippas, Handley, Verde, Roser, McNair, and Evans (2014). *Frontiers in Psychology*, *5*.

Singmann, H., Kellen, D., & Klauer, K. C. (2013). Investigating the other-race effect of germans towards turks and arabs using multinomial processing tree models. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society* (pp. 1330–1335). Cognitive Science Society.

Sison, C. P., & Glaz, J. (1995). Simultaneous Confidence Intervals and Sample Size Determination for Multinomial Proportions. *JASA*, *90*(429), 366–369.

Stephens, R., Dunn, J., & Hayes, B. (2017). A Two-Step Signal Detection Model of Belief Bias. In *Proceedings of the 39th annual meeting of the cognitive science society* (pp. 1138–1143). London,UK: Cognitive Science Society.

Trippas, D., Handley, S. J., & Verde, M. F. (2013). The SDT model of belief bias: Complexity, time, and cognitive ability mediate the effects of believability. *Journal of Experimental Psychology: LMC*, *39*(5), 1393–1402.

Trippas, D., Kellen, D., Singmann, H., Pennycook, G., Koehler, D. J., Fugelsang, J. A., & Dube, C. (in press). Characterizing belief bias in syllogistic reasoning: A hierarchical-bayesian meta-analysis of roc data. *Psychonomic Bulletin & Review*.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the SavageDickey method. *Cognitive Psychology*, *60*(3), 158–189.