How does the latent scope bias occur?: Cognitive modeling for the probabilistic reasoning process of causal explanations under uncertainty

Yuki Tsukamura (tsukka-y@g.ecc.u-tokyo.ac.jp)
Graduate School of Arts and Sciences, The University of Tokyo
3-8-1, Komaba, Meguro-ku, Tokyo 153-8902, Japan.

Taisei Wakai (debatewakai@gmail.com)
Graduate School of Education, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

Asaya Shimojo (asaya.shimojo@konicaminolta.com)
Technology Development Headquarter, Konica Minolta, Inc.
2-7-2 Marunouchi, Chiyoda-ku, Tokyo 100-7015, Japan.

Kazuhiro Ueda (ueda@gregorio.c.u-tokyo.ac.jp)
Graduate School of Arts and Sciences, The University of Tokyo
3-8-1, Komaba, Meguro-ku, Tokyo 153-8902, Japan.

Abstract

When people evaluate explanations in uncertain situations, the *latent scope bias* occurs. It refers to the tendency to perceive explanations that do not include unobservable events as plausible. Previous studies have proposed the *inferred evidence account*, which states that the bias is caused by underestimating the occurrence probability of unobservable events. Additionally, this account assumes that humans use Bayesian probability reasoning in evaluating such explanations. However, previous studies on this bias have not examined the *Bayesian probabilistic reasoning* component. This study measured subjective probabilities of explanations and modeled the reasoning process. As a result, it was found that latent scope bias is caused by Bayesian probabilistic reasoning, compared to the inference using psychological utility. The results also suggest that there are considerable individual differences in the occurrence of latent scope bias. These results support the inferred evidence account. Future studies are required to investigate the factors causing such individual differences.

Keywords: causal explanation; probabilistic reasoning; heuristic; Bayesian cognitive modeling

Introduction

When observing an event, humans try to explain why it happened. This kind of explanation that assumes a cause for a specific event is called *causal explanation* (Lombrozo & Vasilyeva, 2017; Salmon, 1998; hereinafter, *explanation*). In situations where all possible events are observed, previous studies have shown that humans find explanations that can account for more events to be more plausible (Johnson, Johnston, Toig, & Keil, 2014).

However, in the real-world explanations, it is often the case that we cannot observe all the events. For instance, suppose that a patient complains of fever and that there are two possible causes: disease H_N, which causes fever only, and disease H_W, which causes fever and elevated blood glucose level. At this time, the patient may not have had their blood glucose tested and may not know whether their blood glucose level is elevated. We denote that fever is in the *manifest scope* of H_N and H_W, the set of observed events predicted by the explanation, and that elevated blood glucose level is in the *latent scope* of H_W, the set of unobservable events predicted by the explanation.

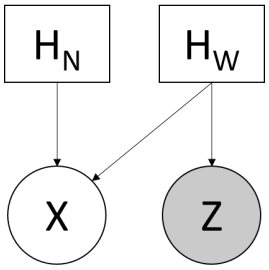

Figure 1: Graphical representation of the causal structure referred to in the Introduction.

In this case, if H_N and H_W occur with equal probability, then normatively, the probability that they caused the patient's fever is equal. On the other hand, it is known that the explanation by H_N, which does not cause the unobservable event, in other words, which has *narrower* latent scope, is preferred to the explanation by H_W (Khemlani, Sussman, & Oppenheimer, 2011; Johnston, Johnson, Koven, & Keil, 2017). This kind of bias is called *latent scope bias*.

Then, why does this bias occur? Johnson, Rajeev-Kumar, and Keil (2016) proposed that this happens because humans underestimate the occurrence probability of unobserved events. This hypothesis is called *inferred evidence account*. The details of this account are described in the following.

Suppose that event X has been observed, and event Z is unobservable. Furthermore, consider that there are two candidates for causes: H_W, which causes both X and Z, and H_N, which causes X but does not cause Z, as shown in Fig. 1. Let I denote the state in which Z is unobservable. Assuming that

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

X and I are conditionally independent given the causes, the following equation (1) can be derived using Bayes rule (for the derivation, see Johnson et al. (2016, Appendix A)):

$$\frac{P(H_N|X,I)}{P(H_W|X,I)} = \frac{P(H_N)}{P(H_W)} \cdot \frac{P(X|H_N)}{P(X|H_W)} \cdot \frac{P(Z|H_N)f^{+Z} + P(-Z|H_N)f^{-Z}}{P(Z|H_W)f^{+Z} + P(-Z|H_W)f^{-Z}} \tag{1}$$

Here, "X" and "Z" denote the states in which the events X and Z are observed, respectively, and "$-Z$" denotes the state in which Z is found absent. In addition, we define $f^{+Z} = P(Z|I)/P(Z), f^{-Z} = P(-Z|I)/P(-Z)$.

In previous studies on latent scope bias, the responses through a two-choice (Johnston et al., 2017; Khemlani et al., 2011) or Likert-type scale (Johnson et al., 2016; Khemlani et al., 2011) were obtained, and there are no studies that measure subjective probability, to the best of our knowledge. However, by obtaining subjective probability responses about the likelihood of explanations, a more detailed investigation of the bias generation process becomes possible by modeling based on equation (1). Specifically, we can directly examine the assumption that "Bayes rule is used to infer the plausibility of explanations, " which is a basis of the inferred evidence account.

In the present study, we obtain subjective probability responses about the likelihood of the explanation, and check whether latent scope bias is occurring in the data. Moreover, we examine the generation process of latent scope bias by fitting possible models and comparing the goodness of fit.

This study is significant in the following manner: First, by formulating the reasoning process based on equation (1) in the form of a probability model, we can provide a quantitative description of the process. Second, by comparing such models, the reasoning process can be specified in a more detailed way. In addition, by estimating the specific value of $P(Z|I)$, it is possible to quantitatively infer the degree of underestimation for each individual. Furthermore, by creating a model that represents the normative reasoning process, we can provide evidence as to whether there is a bias in the reasoning of each individual. The specific hypotheses will be stated after the model is introduced in the next section.

Models

Settings

In this study, we assume the following settings. Assume that $X_1, \ldots, X_m$ is observed, and $Z_1, \ldots, Z_l$ is unobservable. Then, let H_N be the explanation whose manifest scope is $\{X_1, \ldots, X_m\}$ and latent scope is empty, and let H_W be the explanation whose manifest scope is $\{X_1, \ldots, X_m\}$ and latent scope is $\{Z_1, \ldots, Z_l\}$. Here, we tell the participants that the prior probabilities of H_N and H_W are equal and that there is no cause other than H_N and H_W of $X_1, \ldots, X_m, Z_1, \ldots, Z_l$. For the above problem (consider that several such problems were created and numbered using subscript j), we consider obtaining an estimate of the probability of having H_W by participant i and denote the obtained value as $p_{ij}^{(ml)}$. Note that the normative solution in this case is $p_{ij}^{(ml)} = 0.5$.

Based on these settings, we consider how the answer $p_{ij}^{(ml)}$ is generated. In both of the models that follow, we assume that $P(Z|I)$ is underestimated, which is the core assumption of the theory of inferred evidence. The following two types of models can be considered, depending on how the answers are generated.

Bayesian model

One type of the models is based on inferred evidence account. We assume that the events $X_1 \wedge \cdots \wedge X_m$ and $Z_1 \wedge \cdots \wedge Z_l$ are conditionally independent given the causes, H_W always causes $X_1, \ldots, X_m, Z_1, \ldots, Z_l$, and H_N always causes $X_1, \ldots, X_m, -Z_1, \ldots, -Z_l$. Then by expressing the estimated value of $P(Z_1, \ldots, Z_l|I)$ by participant i as a parameter α_i, we can obtain the following equation:

$$\frac{1 - p_{ij}^{(ml)}}{p_{ij}^{(ml)}} = \frac{1 - \alpha_i}{\alpha_i} \tag{2}$$

Taking the logarithm of this and adding the noise according to $N(0, \sigma^2)$, the following model is derived:

$$\log\left(\frac{p_{ij}^{(ml)}}{1 - p_{ij}^{(ml)}}\right) \sim N\left(\log\left(\frac{\alpha_i}{1 - \alpha_i}\right), \sigma^2\right) \tag{3}$$

We call this model *Bayesian model*.

In contrast, $I, Z_1, \ldots, Z_l$ are not conditionally independent, but we can hypothetically think of them as being computed in such ways. Then, we can derive the following model:

$$\log\left(\frac{p_{ij}^{(ml)}}{1 - p_{ij}^{(ml)}}\right) \sim N\left(\log\left(\frac{\alpha_i^l}{(1 - \alpha_i)^l}\right), \sigma^2\right) \tag{4}$$

We call this *independent Bayesian model*.

Utility model

Next, we can consider a model that assumes that responses are made based on expected utility. In other words, this model assumes that psychological utility is obtained depending on the degree of consistency between the events predicted by the explanation and those observed, and that responses are made by comparing the utility of the two explanations.

For each participant i, let u_i be the utility when the event predicted by the cause and the event observed coincide, and let 1 be the utility when they do not coincide. In this case, $U_i(H_W)$, the utility of explanation by H_W, is calculated as fol-

lows:

$$
\begin{aligned}
U_i(H_W) = & \sum_{k=1}^{m} u_i \cdot P(X_k|X_1,\ldots,X_m,I) \\
& + \sum_{k=1}^{l} (u_i \cdot P(Z_k|X_1,\ldots,X_m,I) \qquad (5) \\
& + 1 \cdot P(-Z_k|X_1,\ldots,X_m,I)) \\
= & m \cdot u_i + l \cdot (\alpha_i \dot{u}_i + (1-\alpha_i) \cdot 1) \qquad (6) \\
= & (m + l\alpha_i)u_i + l(1-\alpha_i) \qquad (7)
\end{aligned}
$$

Similarly, the utility of explanation by H_N, $U_i(H_N)$, is computed as follows:

$$U_i(H_N) = (m + l(1-\alpha_i))u_i + l\alpha_i \qquad (8)$$

The process of generating answers from this utility can be considered in several ways. Here, we consider the utility ratio model (defined by Eq. (9)), which takes the ratio of two utilities, and the utility difference model (defined by Eq. (10)), which takes the difference, defined as follows:

$$\log\left(\frac{p_{ij}^{(ml)}}{1-p_{ij}^{(ml)}}\right) \sim N\left(\log\left(\frac{U_i(H_W)}{U_i(H_N)}\right), \sigma^2\right) \qquad (9)$$

$$\log\left(\frac{p_{ij}^{(ml)}}{1-p_{ij}^{(ml)}}\right) \sim N(U_i(H_W) - U_i(H_N), \sigma^2) \qquad (10)$$

Result Prediction

In this study, we first checked whether latent scope bias occurs in subjective probability responses. The latent scope bias is thought to be seen in judgments based on subjective probability, so it is thought that explanations with a narrower latent scope are judged to have higher subjective probability than those with a wider latent scope. Next, based on the inferred evidence account, the Bayesian model is supported in the model comparison. Furthermore, if latent scope bias is explained by the inferred evidence account, the model that takes into account individual differences should be supported, because each individual estimates $P(Z|I)$ independently.

In the present study, we tested whether the above hypotheses are supported or not. In addition to this, we will analyze the quantitative estimates of the parameters to gain insights into the inference process and its individual differences.

Method

Participants

One hundred participants (55 males, 42 females, 3 non-responders; mean age 41.3, $SD = 9.4$), recruited through a crowdsourcing service, engaged in the study via a web browser.

Tasks

The texts about medical diagnosis situations were used for the task, referring to Exp. 2 of Khemlani et al. (2011). To control for the effect of prior knowledge, fictitious disease and substance names were used in the task. The experiment was created using Qualtrics. Participants answered the probability that the patient had one of the two candidate diseases using a 101-point scale from 0% to 100% using an on-screen slider (see Figure 2). They were informed that the prior probabilities of the two candidate diseases were equal and that there was no possibility that there were other candidate diseases.

The tasks were created for the problem settings described above for the following conditions:

$$(m,l) = (0,1),(0,2),(1,1),(1,2),(2,1),(2,2) \qquad (11)$$

Of these, the one with $(m,l) = (1,1),(1,2),(2,1)$ was used because such conditions were set in the previous study (Khemlani et al., 2011), and $(2,2)$ was added from the viewpoint of symmetry. Although there is no previous study on $m = 0$, it was used because it is a special condition that the manifest scope is empty, and it is thought to be effective in considering the condition that the latent scope bias occurs.

Four items were created for each pair (m,l). Two of them asked the participants to answer the probability for the explanation with the wider latent scope among the two explanations presented, and the other two asked the participants to answer for that with the narrower latent scope as inverted scales.

In addition, two items of the Directed Questions Scale (DQS; Maniaci & Rogge, 2014) were created to detect participants who did not respond according to the instructions. In the DQS, participants were presented with the same instructions as in the other tasks, and they were asked to respond with 0 or 100 percent in the end of the instruction.

Procedure

Participants who agreed to participate in the experiment first provided their age and gender. Next, participants were instructed to judge the likelihood that the patient had one disease or the other as a doctor in a fictional world under the condition that the names of the diseases and substances were fictitious. They were also instructed to answer intuitively without thinking too much. In addition, we confirmed the operation of the slider. In this experiment, the slider was pointed at 50% when the screen was displayed, and it was not allowed to proceed to the next screen until the slider was moved. Therefore, even if they wanted to answer 50%, they were instructed to move the slider a little before returning it to 50%.

Next, one practice trial was conducted. In the practice trial, two explanations with different manifest scopes were presented, and the participants were asked to rate the probability of one of them. Subsequently, a total of 26 tasks described above were presented in a randomized order.

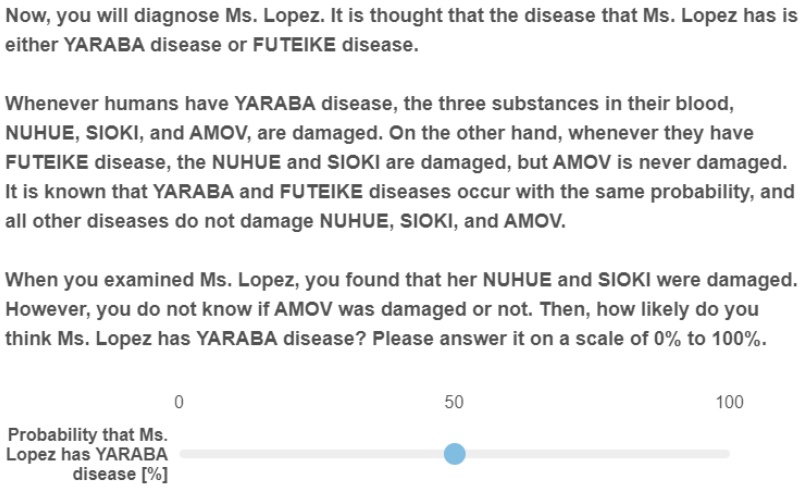

Figure 2: An example of the task, whose parameters are $m = 2, l = 1$.

Table 1: EAP estimate and 95% CI for each μ_m

m	EAP	95% CI
0	0.463	[0.447, 0.479]
1	0.456	[0.444, 0.469]
2	0.460	[0.447, 0.473]

Results

Prior to the analysis, we excluded the data of 21 participants who answered the DQS incorrectly. We also excluded the data of 9 participants who answered 0% or 100% at least once in the main task, due to the divergence of the log odds used in the analysis. As a result, data from 70 participants (39 males, 28 females, and 3 non-responders; mean age 40.1, $SD = 9.2$) were included in the analysis. In addition, reversal items were processed so as to transform all responses into responses $p_{ij}^{(ml)}$ for the explanation with a wider latent scope.

In the following analysis, we employed Bayesian inference using Stan to estimate parameters and compare models. Throughout the analysis, the number of samples per chain was 12,000, of which the warm-up period was 2,000, the number of chains was 4, and the seed value was 4,649.

Bias confirmation

First, we checked whether the latent scope bias was generated by the subjective probability responses. Since $m = 1, 2$ is a condition that has been examined in previous studies, while $m = 0$ is a condition that has not, it is necessary to examine these conditions separately. Therefore, we analyzed the degree of latent scope bias for each m independently. Specifically, we performed beta regression on the preprocessed data as described above. That is, for $m = 0, 1, 2$, we performed Bayesian estimation of the mean of the beta distribution $\mu_m = a_m/(a_m + b_m)$ using the prior distributions $a_m \sim N(0, 5^2), b_m \sim N(0, 5^2), a_m, b_m > 0$ based on the following model:

$$p_{ij}^{(ml)} \sim \mathrm{Beta}(a_m, b_m) \tag{12}$$

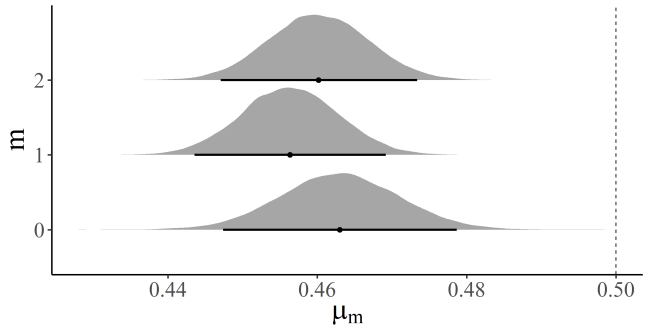

Figure 3: Shape of the posterior distribution of the mean μ_m of the beta distribution, estimated according to the formula (12) for each m of elements in the manifest scope. The dots represent the EAP estimates, the solid horizontal lines represent the 95% CI, and the dashed line represents $\mu_m = 0.5$.

The estimation results are shown in Table 1. The shapes of the posterior distributions, expected a posteriori (EAP) estimates, and their 95% credible intervals (CI) are shown in Figure 3. Since the upper bound of the 95% CI was < 0.5 in all cases, we can say that the latent scope bias occurred even in responses based on subjective probability. It is also found that the bias occurred in the $m = 0$ condition. In contrast, μ_m was close to 0.5, suggesting that the effect of latent scope bias was not so large.

Model comparison

Next, for each model, we calculated the goodness of fit for the obtained data. In estimating the parameters, the prior distributions were set as $\alpha \sim U(0, 1), u \sim N(1, 1^2)(u > 1), \sigma \sim$

Table 2: Calculated WAIC and FE for each model.

	WAIC	FE
Bayesian model	1.097	1947.6
independent Bayesian model	1.118	2008.7
Utility ratio model	1.140	1998.9
Utility difference model	1.113	1971.8

Table 3: Calculated WAIC and FE for the model with and without individual differences.

	WAIC	FE
model with individual differences	1.097	1947.6
model without individual differences	1.203	2027.1

$N(0,5^2)(\sigma > 0)$. Widely applicable information criterion (WAIC) and free energy (FE) were computed as measures of goodness of fit. To calculate the free energy, we used the bridgesampling package in R.

The results are shown in Table 2. Since the Bayesian model had the lowest values both for WAIC and for FE, the Bayesian model fit the data best and was thought to best represent the participants' reasoning process. For this reason, we decided to use the Bayesian model in subsequent analyses.

In addition, to examine individual differences of the parameters, we created a model in which α in the Bayesian model is a common parameter for all participants. The model created here is called the *model without individual differences*, and the original model is called the *model with individual differences*.

A comparison of the goodness of fit for these models is shown in Table 3. Both the WAIC and FE were lower in the model with individual differences. This result strongly supports the model with individual differences, suggesting that the individual differences in α are considerable.

The mean and 95% CI of the posterior distribution of α_i for each individual are shown in Figure 4. There were 18 participants whose 95% CI for α_i did not include 0.5. Of these, 17 had upper bounds on the confidence interval < 0.5. Therefore, it was suggested that a certain number of participants underestimated $P(Z|I)$. In contrast, for 52 participants (over 70% of the total), the 95% CI included 0.5, and for one participant, the lower bound of the 95% confidence interval was > 0.5. This suggests that many participants did not necessarily underestimate $P(Z|I)$.

Comparison with a normative model

The previous analysis suggests that there were some individual differences in α_i. In contrast, it also suggests that the underestimation, i.e., the latent scope bias, did not necessarily occur in many participants. In particular, participants around $\alpha_i = 0.5$ did not exhibit latent scope bias, suggesting that they were reasoning normatively. However, inference with credible intervals cannot provide positive evidence that each individual is doing so. Therefore, we created a normative response model and analyzed which of the Bayes or normative models would fit the data for each participant.

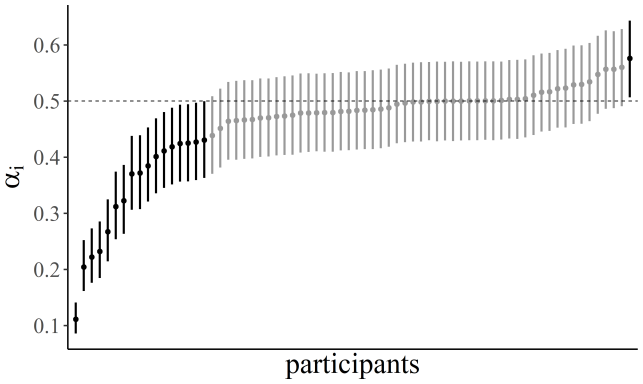

Figure 4: Estimation results of α_i for each participant in the Bayesian model, in ascending order. The dots represent the EAP estimates, the solid lines represent the 95% CIs, and the dashed line represents $\alpha_i = 0.5$. Participants whose 95% CI does not include 0.5 are drawn in darker colors.

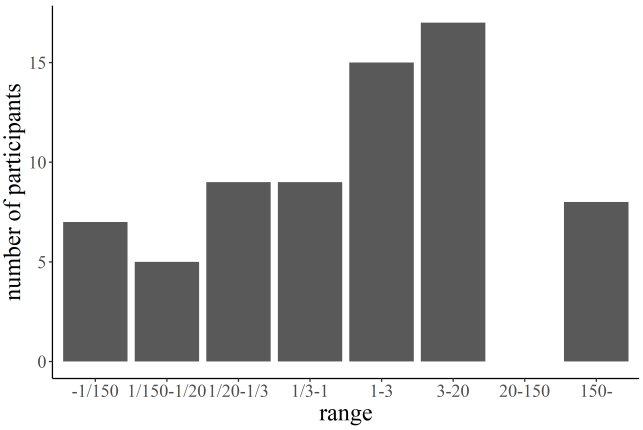

Figure 5: Histogram of BF_i, classified based on Kass and Raftery (1995).

In the case of a normative response, the log odds of $p_{ij}^{(ml)}$ are expected to follow a normal distribution with mean 0. In other words, the following model is called the *normative model*:

$$\log\left(\frac{p_{ij}^{(ml)}}{1-p_{ij}^{(ml)}}\right) \sim N(0,\sigma^2) \tag{13}$$

For each participant, the Bayes factor BF_i was calculated for the obtained data y_i. BF_i is defined by the following equation, i.e., the ratio of marginal likelihoods of the normative model (denoted as model M_0) and the Bayesian model (denoted as model M_1):

$$BF_i = \frac{p(y_i|M_1)}{p(y_i|M_0)} \tag{14}$$

Note that in calculating BF_i for each individual, the population-level parameter σ of each model was fixed to the EAP estimated using everyone's data.

The histogram of the calculated BF_i is shown in Figure 5. In this figure, Bayes factors were classified based on the criterion of Kass and Raftery (1995) [1]. Twenty-one of the 70 participants supported the normative model at the "positive" level or higher, while 25 participants supported the Bayesian model at the "positive" level or higher. This suggests that although some participants showed latent scope bias, the bias did not appear in a certain number of participants. Furthermore, for each model, there were several participants who had "very strong" levels of support, suggesting that there were large individual differences in latent scope bias.

Discussion

This study focused on the inferred evidence account as the process of the latent scope bias, and directly verified the Bayesian probabilistic reasoning component by cognitive modeling using the subjective probability response data. The bias confirmation showed that latent scope bias was generated even in responses based on subjective probability. In addition, the results of the model comparison suggested that the inferred evidence account proposed by Johnson et al. (2016) was a relatively reasonable explanation. The analysis of individual differences suggested that there are certain individual differences in the occurrence of latent scope bias.

Furthermore, the results of the beta regression (Figure 3) suggest that the effect size of the latent scope bias is not very large, and the Bayes factor (Figure 5) suggests that although there were indeed a certain number of participants who had a bias, but at least for the participants in this experiment, there were not many who experienced the latent scope bias.

In this study, it is suggested that it is more appropriate to assume that we are reasoning with probabilities, rather than reasoning with utilities, as an inference process for the likelihood of explanations. Together with the result that the biased term in the (1) equation is $P(Z|I)$ (Johnson et al., 2016), this result supports the theory of inferred evidence as the process of generating latent scope bias. However, it should be noted that the results of the model comparison in this study are only within the scope of the models examined, and different results may be obtained by considering other models.

In previous studies, individual differences in latent scope bias or the heuristics that cause it have not been examined. The present study quantitatively showed that some individuals engage in normative reasoning, while others do not, suggesting that the majority of individuals engage in normative reasoning. In future research, it should be considered that latent scope bias does not necessarily occur for everyone, and future research should consider factors causing such individual differences.

Future research is to examine whether this study's findings can be generalized to other types of responses (e.g., Likert-type scale). Additionally, the model used in this study assumes that the participants understand the questions correctly. However, it is possible that the participants do not understand the question statement according to normative probability theory, as seen in the criticisms of the research on representativeness heuristics (e.g., Hertwig & Gigerenzer, 1999; Hertwig, Benz & Krauss, 2008). Further studies are needed to address this point as well.

[1] *Note*: $1 \leq BF_i \leq 3$: Not worth more than a bare mention, $3 \leq BF_i \leq 20$: positive evidence for M_1, $20 \leq BF_i \leq 150$: strong evidence for M_1, $BF_i \geq 150$: very strong evidence for M_1

Conclusion

In this study, we modeled how the latent scope bias arises, and the results showed that the Bayesian inference process proposed in previous studies was the most supported. Moreover, the results suggest that there are individual differences in the occurrence of latent scope bias and that there are some individuals who have bias, whereas others show normative inference.

Acknowledgments

This study was supported by JST CREST JP-MJCR19A1 and JSPS KAKENHI JP-22H03911.

References

Hertwig, R., Benz, B., & Krauss, S. (2008). The conjunction fallacy and the many meanings of *and*. *Cognition*, *108*(3), 740–753.

Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*(4), 275–305.

Johnson, S. G. B., Johnston, A. M., Toig, A. E., & Keil, F. C. (2014). Explanatory scope informs causal strength inferences. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual meeting of the cognitive science society* (pp. 2453–2458). Austin, TX: Cognitive Science Society.

Johnson, S. G. B., Rajeev-Kumar, G., & Keil, F. C. (2016). Sense-making under ignorance. *Cognitive Psychology*, *89*, 39–70.

Johnston, A. M., Johnson, S. G. B., Koven, M. L., & Keil, F. C. (2017). Little bayesians or little einsteins? probability and explanatory virtue in children's inferences. *Developmental Science*, *20*(6), e12483.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Khemlani, S. S., Sussman, A. B., & Oppenheimer, D. M. (2011). Harry potter and the sorcerer's scope: latent scope biases in explanatory reasoning. *Memory and Cognition*, *39*(3), 527–535.

Lombrozo, T., & Vasilyeva, N. (2017). Causal explanation. In M. R. Waldmann (Ed.), *The oxford handbook of causal reasoning* (pp. 415–432). New York: Oxford University Press.

Salmon, W. (1998). *Causality and explanation*. New York: Oxford University Press.