Symmetric Bias in Reasoning: Error Analysis of Indeterminate Term Series Problems

Takafumi Aoi (s2340002@jaist.ac.jp)
Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa, Japan

Shohei Hidaka (shhidaka@jaist.ac.jp)
Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa, Japan

Abstract

In term series problems where multiple mental models can be constructed, partial-order models can be created as mental representations, which make it easier to perceive the symmetry of the terms. To test these hypotheses, we categorized multi-model (indeterminate) term series problems according to the patterns of partial-order models that could be constructed, and analyzed the reasoning performance for each pattern. These results suggest that reasoners tend to use the symmetry of terms to reduce the cognitive load of reasoning. Analysis of the patterns of incorrect answers also suggests that attempts to exploit the symmetry of the term may be biased, leading to errors in reasoning.

Keywords: mental model; term series problem; annotation; symmetry bias; preferred model; partial-order model

Introduction

The term series problem is the problem of reasoning from the premises that "A is larger than B" and "B is larger than C" to "A is larger than C," "A is the largest," and so on. In some studies, this is called a linear syllogism or relational reasoning. The term series problem has been the subject of many reasoning studies because it is the simplest category of deductive reasoning problems and the variables can be easily manipulated in experiments.

De Soto, London, and Handel (1965) and Huttenlocher (1968) reported that reasoners working on the term series problem construct spatial representations in which terms are arranged according to the order described in the premises. For example, given the premises "A is larger than B" and "B is larger than C," the representation ABC is constructed (where the terms on the left are larger; if the term on the right is larger, it becomes CBA). The idea that such spatial representations are constructed during reasoning was subsequently embraced by mental model theory (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991).

Term series problems are problems in which more than one model can be constructed. For example, given the premises that "A is larger than B" and "A is larger than C," two models can be constructed, ABC and ACB, because it is not known whether B or C is larger. Such problems are called multiple-model problems (because more than one model can be constructed) or indeterminate problems (because the order of all the terms is not fixed to one).

It is worth remarking that the term "model" in the above line of studies has been almost exclusively meant to express a *total-order model*. As all term series problems consider some kind of *order* relationship among the terms, we briefly summarize the mathematical notions of order. One of the minimally axiomized notions of order is *preorder*. In any preordered set with an order operator $\geq$, it holds (1) *reflexivity:* $x \geq x$ and (2) *transitivity*: $x \geq z$ if $x \geq y$ and $y \geq z$ for any members $x, y,$ and z in it. *Partial-order* is a restricted order property on top of the preorder. In any partial-order set with an order operator $\geq$, it holds (1) *reflexivity*, (2) *transitivity*, and (3) *antisymmetry*: $x = y$, if $x \geq y$ and $y \geq x$. *Total-order* is further restricted: In any total-order set with an order operator $\geq$, it holds (1) *reflexivity*, (2) *transitivity*, (3) *antisymmetry* and (4) *totality*: $x \geq y$ or $y \geq x$ holds for any pair of the members x and y in it.

Previous studies on term series problems have implicitly assumed that participants build their mental models in the form of total-order. This is because there are "multiple mental models" for what they call indeterminate problems, while in fact there is one and no more than two models in the form of *partial-order* for the indeterminate problems. Namely, the count of "models" may depend on which class of order relationships is (implicitly or explicitly) assumed in each study. Thus, "indeterminate term series problem" can be defined by any term series problems with a set of assumptions that admits *multiple total-order* mental models, while "determinate term series problem" is then defined by that which admits only a *single total-order* mental model.

With this in mind, we review previous studies on indeterminate term series problems as follows. Early research predicted that the difficulty of indeterminate model problems would be proportional to the number of total-order models that could be constructed (Byrne & Johnson-Laird, 1989). However, subsequent research found that there was no correlation between reasoning performance and the number of total-order models in indeterminate problems (Knauff, Rauh, Schlieder & Strube, 1998; Vandierendonck, De Vooght, Desimpelaere & Dierckx, 1999). Based on these results, the researchers concluded that not all total-order models are constructed in solving the indeterminate term series problem. Rather, only one or a few total-order models are constructed, and some additional mental models may be constructed, if required.

Related Research

Preferred model theory has been proposed as a theory of cognitive processes in the indeterminate problem. According to this theory, one preferred total-order model should be constructed as a mental representation for solving an indeterminate problem. Jahn, Knauff and Johnson-Laird (2005) analyzed responses to indeterminate term series problems and found that when multiple terms

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

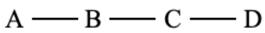
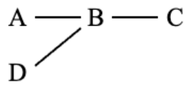
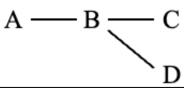
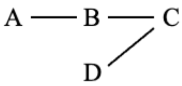
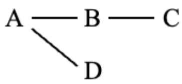
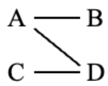
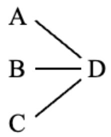
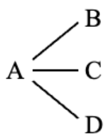
are equally appropriate for a given question, the term appearing in the earlier read premise tends to be placed next to the referred term in the model, whereas those appearing in the later read premise tend to be placed outside. For example, if the premises are read in the order "A is to the left of C" and "B is to the left of C," then of the two terms (A and B) to the left of the referred term C, A, which is read first, is placed immediately to the left of C, and B is placed beyond it, creating the model BAC. This principle later became known as the *first free fit principle* (*fff-principle*). Rauh et al. (2005) argued that alternative models are created by revising parts of the preferred model. However, if the model revision is done incorrectly, the model can become inconsistent with the given premises. Such errors can be prevented by *annotations* added to a preferred model. The role of annotations is to instruct the reasoner as to which parts of the model can be changed and to what extent. For example, in the example above, the annotation "to the left of C" is given to the term B, which was not allowed to be adjacent to the referred term C. The annotation shows that B can move anywhere if it is to the left of C; therefore, an alternative model ABC can be constructed by replacing B with A in the preferred model BAC, whereas model ACB cannot be constructed because B cannot be placed to the right of C.

The PRISM proposed by Ragni and Knauff (2013) is a computational model that constructs preferred models with annotations following the fff-principle in spatial relational reasoning problems. PRISM creates a preferred model by placing terms according to the fff-principle on a two-dimensional spatial array that simulates spatial working memory. In addition, PRISM generates alternative models by repeatedly swapping the positions of annotated terms and their adjacent terms in a model, starting with a preferred model. An important parameter in PRISM is the number of *focus operations*; the focus moves one cell at a time on the spatial array, and operations such as inserting terms are performed on the cell where the focus is located. The number of focus operations increases each time the focus moves or an operation such as term insertion is performed. In PRISM, the number of focus operations is a measure that explains the difficulty of term series problems and why indeterminate problems are more difficult than determinate problems.

Apart from annotated preferred models such as those implemented in PRISM, another alternative form of mental representation in the indeterminate problem is the partial-order model, in which the order relationship is expressed by a *tree-like* network rather than a linear sequence. For example, in Table 1, 2M-FS shows the partial-order models that can be constructed from the premises that "A is larger than B," "B is larger than C," and "B is larger than D."

A class of order relations, where any pair of elements needs to have some order relation, is called total-order. Another class of order relations that accepts some pair of elements without any order relationship is called partial-order. The preferred models essentially assume a total-order relationship in mental representation, whereas the partial-order model admits not only total-order but also partial-order models. In Table 1, the total-order models that can be constructed from the respective partial-order models are included. For example, in 2M-FS of Table 1, D can be inserted anywhere to the right of B; thus, ABDC can be constructed by inserting it between B and C, and ABCD can be constructed by inserting it to the right of C.

Table 1: All possible partial-order models that satisfy some term series problem with three premises and four terms (the more the term is on the left side, the larger the term is). A line between terms indicates that the relationship between them is described in premises. Relationships between terms not connected by a line are unknown. Each type is named using the number of total-order models and the direction of fork. For example, "3M" in 3M-FL means that three total-order models can be created, and "FL" means that there is a fork on the larger side. It becomes "FS" if there is a fork on the smaller side.

Type	Total-order models	Partial-order model
1M	ABCD	
2M-FL	ADBC DABC	
2M-FS	ABCD ABDC	
3M-FL	ABDC ADBC DABC	
3M-FS	ADBC ABDC ABCD	
5M	ACBD ACDB ABCD CABD CADB	
6M-FL	ABCD ACBD BACD BCAD CABD CBAD	
6M-FS	ABCD ABDC ACBD ACDB ADBC ADCB	

There are reports supporting partial-order models (note that in these previous studies, the partial-order model is referred to as the *isomeric model*). Schaeken, Van der Henst and Schroyens (2007) allowed participants to use a paper and pencil when solving indeterminate term series problems. One group of participants drew partial-order models and performed better than the other group, which did not draw them. The authors considered this result as evidence that reasoners construct a partial-order model in their mental representation. Vandierendonck, Dierckx, and De Vooght (2004) also reported that their experimental results could be better explained by considering that reasoners do not immediately construct partial-order models when reading the premises, but first construct an annotated model and then elaborate it to construct a partial-order model when necessary.

Hypothesis

As aforementioned, in the indeterminate term series problem, the idea that reasoners construct models with annotations or forks has been favored in previous research. Of the ideas in previous studies, we consider the partial-order model to be promising, because it is easier to perceive the symmetry of the term pairs in the partial-order model. If there is a pair of terms, such as C and D in 2M-FS in Table 1, such that the model is compatible with the premise even if one term and the other are swapped, then the terms are defined as symmetric. If we recognize that C and D are symmetric, we can easily see that they are both the third or fourth-largest, and that ABCD and ABDC can be constructed as total-order models. It is believed that it is possible to efficiently reason problems with symmetric term pairs by recognizing the symmetry of the terms. However, for problems without symmetric term pairs, such as in 3M-FS in Table 1, this efficiency is not possible.

We hypothesize that reasoners construct partial-order models as their mental representations and that the difficulty of term series problems is correlated with the structure of the partial-order models to be constructed. To explain our hypothesis in detail, it is useful to categorize term series problems according to the type of partial-order model that satisfies given premises. We can test our hypotheses by analyzing the correlation between the type of partial-model and reasoning performance. In this study, we focus on a class of term series problems with four terms and three premises, because such a class of problems offers a minimal set of a wide variety of partial-order models. Eight possible partial-order models satisfy four-term series problems with three premises and four terms (Table 1). Each graph in Table 1 depicts a partial-order model with the terms A, B, C, and D, and the series of terms below each graph (e.g., ABCD) show the list of possible total-order model(s) satisfying the corresponding problem. Each type of problem corresponds to a unique partial-order model, but has one or more total-order models.

One may find some symmetric structures in some partial-order models. For example, the 2M-FL has the term A and D exchangeable—that is, exchanging A and D does not affect the graph essentially. Likewise, C and D in 2M-FS; A, B, and C in 6M-FL; and B, C, and D in 6M-FS are mutually exchangeable. These types of partial-order models are called *symmetric*—their graph structures are invariant under the exchange of a particular subset of vertices. The other types are asymmetric. For example, exchanging any of the four terms in 3M-FL would create another partial-order model that does not follow the same premises.

We hypothesize that this symmetry structure in partial-order models plays a role in the reasoning of term series problems, as symmetry in partial-order models reduces the cognitive load in reasoning. In other words, the reasoner does not need to accurately identify and locate every term in the problem if some pairs of terms in the partial-order models are exchangeable (i.e., symmetric). This hypothesis also predicts that the problems with 3M-FL and 3M-FS are more difficult than those with 2M-FL and 2M-FS, because 3M-FL and 3M-FS look "symmetric" but are actually not. If the reasoner employs reasoning strategies exploiting the symmetric structure of partial-order models, 3M-FL and 3M-FS can be rather confusing, as their graph shapes are similar but not as symmetric as 2M-FL and 2M-FS, respectively.

Experiment

According to the symmetry hypothesis, problems with symmetric term pairs (2M-FL, 2M-FS, 6M-FL, and 6M-FS) are easier to solve than their asymmetric counterparts (3M-FL, 3M-FS, and 5M), resulting in higher correct response rates and shorter average correct response times. To test this hypothesis, we manipulated the types of partial-order models and the place of the term to be questioned. There are 8 graphs (Table 1) × 4 term positions = 32 patterns of term series problems were used in our experiment. In our term series problems, the participants were asked to answer all the terms that possibly take the k-th largest place among the four terms. For example, the correct answer to the second place in a problem with 3M-FL is "A, B, and D" (Table 1).

The results of this experiment were analyzed using the correct response rate and average response time for each pattern, treating only those cases in which all terms were selected without excess or deficiency as correct (for response times, only data from correct responses were considered and averaged). Participants were recruited online to participate in the experiment. The experimental code was created using PsychoPy, open-source software for creating psychological experiments, and was run on the Pavlovia server service (https://pavlovia.org/). This allowed anyone with access to the URL provided by Pavlovia to participate in the experiment online.

Participants

The 80 participants were recruited using CrowdWorks (http://crowdworks.jp), a Japanese crowdsourcing service (applicants had to be Japanese speakers between the ages of 20 and 65 years). One of the participants spent an unusually long time answering a question (just under 45 min); therefore, this piece of data was excluded from the analysis because it was considered a source of noise.

Procedure

The participants were informed of the URL of the Pavlovia site and they began the experiment by navigating to the site.

The experimental screen displayed all three premises and one question simultaneously. The three premises were arranged one sentence at a time from top to bottom, and below them was a question text that asked, "What is the k-th largest possibility? (Please select all)," where k = 1, 2, 3, or 4. The four checkboxes were arranged horizontally to select the terms that needed to be answered. Below that was a button for submitting the answer. The setup ensured that participants could not accidentally skip to the next question. The time from when the question screen was displayed to when the "Answer" button was pressed was recorded as the response time for the question.

All premises were in the form "A is larger than B"; the opposite phrase "smaller than" was not used in any trial. The letters A, B, C, and D were used as terms in the premises. The format of the questions was designed such that the position of the term and the order of the premises changed randomly from time to time.

Block Design

Each participant solved 75 term series problems, consisting of practice, fixed, and random blocks of 3, 40, and 32 questions, respectively. The practice block was intended to familiarize the participants with the experimental procedure, so that every participant was aware that there could be more than one term as the answer. In the fixed block, each of the five consecutive questions asked about the same partial-order model with the same set of premises, and the place (rank) of the terms to be answered only varied across these consecutive questions. As this process was conducted for all eight partial-order models, the number of questions in this block was 40 (5 × 8). The five consecutive questions were asked randomly, with the restriction that the first question was asked about any first to fourth rank, and the second through fifth questions asked about all ranks from 1 to 4 within those four questions. The fixed block was expected to extract the time required to search the model because only the rank to be answered changed, and the model construction was supposed to be completed in the first question as long as it was memorized through the remaining four questions.

In the random block, each of the 32 patterns (4 places of terms to be answered x 8 partial-order models) were presented once in a randomized order. As participants had already experienced each question pattern once in the fixed block, the random block was expected to produce results without additional effort owing to unfamiliarity, which can occur when questions are new to participants.

At the end of the experiment, participants were directed to Google Forms to answer whether they had used writing utensils during the experiment. Of the 79 participants, 24 did not use a writing utensils at all, while the remaining 55 used a writing utensils for at least one question. The experimental results and discussions described below did not differ significantly between the two groups, so this paper does not delve into the impact of writing utensils usage.

Result

Analysis of Reasoning Performance

The percentage of correct answers for each of the 32-pattern questions in the random block and the average response times for correct answers are shown in Figure 1. As predicted, correct answer ratios were higher and correct response times were shorter for questions where the term pairs were symmetric (2M-FL, 2M-FS, 6M-FL, and 6M-FS) than for those where they were not (3M-FL, 3M-FS, and 5M). This trend was also observed for the fixed block (not shown). This suggests that participants reasoned efficiently with term symmetry, supporting this hypothesis.

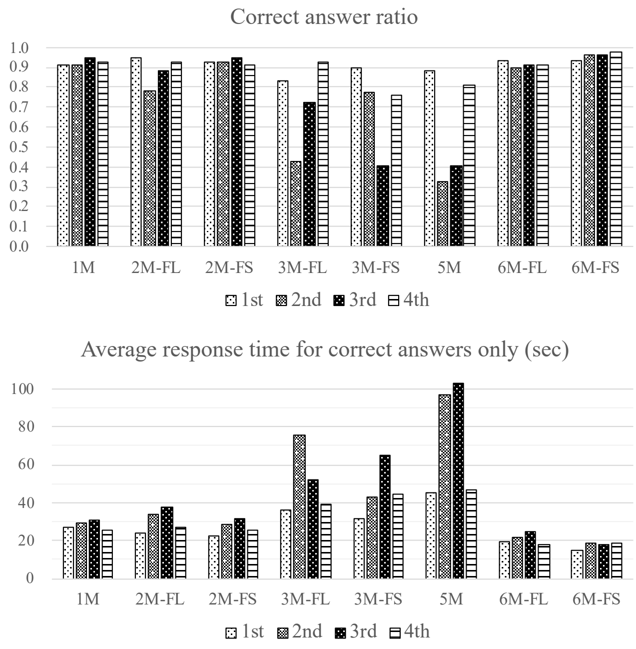


Figure 1: Experimental result of four-term series problems.

Preferred Model and Symmetry Hypothesis

We also analyzed whether the preferred model theory could account for this result, but we found this unlikely, as follows: Consider PRISM, a simulation program that implements the preferred model theory to solve the question of all the largest possible terms in 6M-FL. The correct answers are all three, A, B, and C. To answer this question correctly using PRISM, one would have to first construct a preferred model, then modify that preferred model twice to generate two alternative models, and construct three total-order models. As this process requires many focus operations, one would expect the time required to answer correctly to be longer. In fact, the average correct response time for 6M-FL is equal to or shorter than the average correct response time for 1M, which does not require the construction of an alternative model.

However, this result can be explained if we consider that the participants constructed the partial-order model and recognized that the three terms A, B, and C on the side larger than D are symmetric to each other. Using the symmetry of the terms, one can immediately see that D, which is on the side smaller than the three terms, is the fourth-largest and that the possible ranks of A, B, and C are first, second, or third.

To numerically analyze the factors that contribute to reasoning performance, a logistic regression was run to predict correct (set to 1) and incorrect answers (set to 0). For each question, the features that seemed likely to contribute to the performance were extracted from the structure of the partial-order models. The features defined by focus operations in PRISM were extracted as candidate explanatory variables, and features with high correlation were removed to avoid multicollinearity; 12 features remained. Logistic regression with an intercept was performed on these features, and the standardized partial regression coefficients for each feature, as well as the p-values and t-values for the t-test with the null hypothesis that the standardized partial regression coefficient was zero, were also obtained. An explanation of each feature and the results of the logistic regression (standardized partial regression coefficients, t-values, and p-values) are presented in Table 2. The standardized partial regression coefficients for the features that were significant ($p < 0.05$) were (e), (f), (g), (h), (k), and (l). None of the effects of the three features extracted from PRISM ((a), (b), and (c)) were significant.

Although it is difficult to fully explain all these results theoretically, we believe that there is an explanation for the negative value of (k) and the positive value of (l). As the number of answer terms ((k)) increases, it becomes more difficult to answer them without over- or under-answering, which has a negative impact on the ease of answering them correctly. However, even if the number of correct terms is large, if they are symmetric (i.e., if (l) is 1), the difficulty decreases. For example, the question asking for the second-largest term in 3M-FL and the question asking for the first-largest term in 6M-FL both had three correct answers, but the percentage of correct answers in the latter was approximately twice that in the former. This may be because the latter question can be efficiently answered by exploiting the symmetry of the three correct answer terms.

Analysis of Error Patterns

The symmetry hypothesis predicts a particular type of error in reasoning for indeterminate term series problems. In other words, reasoning on asymmetric but partially symmetric partial-order models such as 3M-FS, 3M-FL, and 5L induces more error responses toward symmetric-biased patterns. Therefore, we examined the error patterns of term series problems with less than 50% correct questions. Specifically, the error patterns for the four questions were analyzed: the question about the second-largest term in 3M-FL, the third-largest term in 3M-FS, and the second- and third-largest terms in 5M. Table 3 shows the most frequent incorrect answer patterns committed by participants for each of the four question types, along with the percentage of that pattern among all incorrect answers for each question. Incorrect answers to each question showed a bias toward certain patterns. For all four types of questions, the correct answers were subsets of the three terms, and approximately half the incorrect answers missed one specific term, answering only B and D in the question asking for the second-largest term in 3M-FL, B and D in the question asking for the third-largest term in 3M-FS, A and C in the question asking for the second-largest term in 5M, and B and D in the question asking for the third-largest term in 5M. This trend was the same for both the random and fixed blocks (the ratios shown in Table 3 are the result of mixing both blocks).

We found that the symmetry(-like) structure in the partial-order model was again the key to explaining the bias in the error patterns. Note that the pattern that accounted for approximately half the incorrect answers to each of the four types of questions was the pattern that answered only two terms with identical horizontal axis coordinates in the partial-order model. In 2M-FL, 2M-FS, 6M-FL, and 6M-FS, terms with the same horizontal axis coordinates were symmetric, but not in 3M-FL, 3M-FS, and 5M. However, it is possible that many participants misinterpreted these terms as symmetric. In this case, a partial-order model was created with lines drawn between AD in 3M-FL, CD in 3M-FS, and BC in 5M, which should not exist. Under this illusory partial-order model, it would be correct to answer B and D to the question asking for the second-largest term in 3M-FL and the third-largest term in 3M-FS, A and C to the question asking for the second-largest term in 5M, and B and D to the question asking for the third-largest term in 5M. Even if an illusory model was constructed, the correct answers to the questions asking for the third- and fourth-largest terms of 3M-FL, the first- and second-largest terms of 3M-FS, and the first- and fourth-largest terms of 5M would not change. In contrast, the correct answers changed for questions that asked for the first-largest term in 3M-FL

Table 2: Standardized partial regression coefficients, t-value, and p-value derived from logistic regression for correct answer ratio. The asterisk at each letter with parentheses indicate the variable has its coefficient significantly larger or smaller than zero (p-value < 0.05).

	Explanation of features	Coefficient	t-value	p-value
(a)	the number of times the focus moves through a cell in PRISM	-0.03	-0.10	0.917
(b)	the number of times the focus places a term in a cell in PRISM	0.03	0.25	0.801
(c)	the number of times the focus changes the direction of movement in PRISM	-0.13	-0.89	0.372
(d)	the number of all possible total-order models that can be constructed	-0.65	-1.32	0.185
(e)*	1 if the fork is on the larger side in the partial-order model, 0 otherwise	-1.27	-4.15	< .001
(f)*	1 if the fork is on the smaller side in the partial-order model, 0 otherwise	-1.08	-3.61	< .001
(g)*	the maximum number of terms that can be connected without branching in the partial-order model	-1.91	-2.75	0.006
(h)*	1 if the question asks for the first-largest term, 0 otherwise	0.23	2.70	0.007
(i)	1 if the question asks for the second-largest term, 0 otherwise	-0.07	-1.03	0.301
(j)	1 if the question asks for the fourth-largest term, 0 otherwise	0.13	1.58	0.113
(k)*	Number of correct answers	-0.83	-7.06	< .001
(l)*	1 if more than one term is correct answer and they are symmetric, 0 otherwise	0.39	2.61	0.009

and the fourth-largest term in 3M-FS. Analysis of the error patterns in these questions showed that approximately half the incorrect answers in the first-largest term question in 3M-FL were answered with only A, and approximately half the incorrect answers in the fourth-largest term question in 3M-FS were answered with only C. Both patterns are consistent with the illusion model. Based on the results of the present study, it is possible that the misidentification of non-symmetric terms as symmetric may have biased the error patterns.

Table 3: For each question where the correct answer ratio was less than 50%, the top three most frequently occurring incorrect answer patterns and the proportion of those patterns to the number of incorrect answers in each question.

Problem	Correct answer	Incorrect answer	Proportion
3M-FL 2nd-largest	A, B, D	B, D	57%
		A, D	20%
		B	8%
3M-FS 3rd-largest	B, C, D	B, D	45%
		C, D	36%
		C	6%
5M 2nd-largest	A, B, C	A, C	54%
		B, C	15%
		A	5%
5M 3rd-largest	B, C, D	B, D	61%
		B, C	17%
		D	4%

We examine whether PRISM can explain this bias in error patterns. For example, in the question asking for the second-largest term in 3M-FL, approximately 50% of the incorrect answers listed B and D, missing the presence of A. Here, in about a third of the cases where the incorrect answers were B and D, the first two premises were "A is larger than B" and "B is larger than C," and the third and final premise was "D is larger than C." In such a case, according to PRISM, the first two premises construct a model ABC, and when the last premise is read, a preferred model DABC is created with the annotation "larger than C" assigned to D. As the second-largest term in this preferred model is A, it is unlikely that reasoners would overlook A when asked about the second-largest term. Similarly, in the 3M-FS question asking for the third-largest term, approximately 50% of the incorrect answers listing B and D overlooked the presence of C. In about one-third of the cases where only B and D were answered, the first two premises were "A is larger than B" and "B is larger than C," and the third and final premise was "A is larger than D." In this case, PRISM would construct a preferred model ABCD with D annotated as "smaller than A," so it is unlikely that C would be overlooked when asked for the third-largest term. As the three premises were presented simultaneously in this experiment, participants did not necessarily read the premises in the order in which they were arranged. Perhaps participants read the premises in the order in which they were most comfortable, which would be in 3M-FL and 3M-FS, where the first two premises are "A is larger than B" and "B is larger than C." This order did not cause indeterminacy until the third and final premises were read. The reading order is the same as that described above. Therefore, if we assume that participants read the premises in the order in which they were arranged or in the order in which they felt comfortable, PRISM is unlikely to reproduce the error pattern bias observed in this experiment.

General Discussion

The purpose of this study was to investigate the class of mental model reasoning on indeterminate term series problem constructs, annotated total-order models or partial-order models. In previous studies, these two classes of models supported distinct sets of empirical studies. In this study, we hypothesized that partial-order models are likely to be employed in reasoning indeterminate term series problems, and proposed a strategic use of symmetry in partial-order models. As this hypothesized symmetry is generally available only in partial-order models and not in total-order models with or without annotations, the symmetry hypothesis would exclusively assume partial-order models.

To test our hypothesis, we investigated reasoning behaviors on a minimal set of term series problems with several variations in partial-order models and analyzed the correct ratio and error rates.

Our analysis suggested that the symmetry hypothesis was consistent with the results of the present experiment; however, it would be difficult to explain it with annotated total-order models such as PRISM.

According to Goodwin and Johnson-Laird (2005), reasoners develop different strategies to make their reasoning more efficient. These strategies reflect the nature of the problem they are working on. The questions in this experiment were in a format that asked participants to answer all terms that could be assigned a particular rank. It is possible that the format of the questions forced the participants to develop strategies for constructing partial-order models. In other words, if the same format as in previous research (e.g., asking participants to judge the validity of the description of the relationship between certain terms or simply asking them to answer all possible models that could be constructed) was used, it is possible that inferences could have been made using a procedure such as PRISM. However, even in the fixed block, in which participants may not have had time to become familiar with the partial-order model, the percentage correct, average correct response time, and error patterns followed the same trend as in the random block. This suggests that the reasoning strategy using the term symmetry does not require familiarization with or learning of mental models suitable for a given set of premises. In this sense, symmetry use in indeterminate term series problems is likely to be a more general strategy or a cognitive bias that may enhance human reasoning performance in certain situations.

One adaptive aspect of this "symmetry bias" in term series reasoning is to reduce cognitive load. This symmetry bias may have allowed more efficient reasoning for problems with smaller working memory if the partial-order models were symmetrical.

Acknowledgments

This work is supported by JSPS KAKENHI JP 20H04994, JP23H0369, JST PRESTO JP-MJPR20C9, Japan. We would like to thank Editage (www.editage.jp) for proofreading an early version of this manuscript.

References

Byrne, R., & Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory and Language, 28,* 564 – 575.

De Soto, C. B., London, M., & Handel, S. (1965). Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology, 2,* 513–521.

Goodwin, G. P., & Johnson-Laird, P. N. (2005). Reasoning about relations. *Psychological Review, 112,* 468–493.

Huttenlocher, J. (1968). Constructing spatial images: A strategy in reasoning. *Psychological Review, 75,* 550–560.

Jahn, G., Knauff, M., & Johnson-Laird, P. (2005). Preferred mental models in spatial reasoning. *Proceedings of the twenty-seventh annual conference of the cognitive science society* (pp. 1036-1041). Mahwah, NJ: Erlbaum.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge, MA: Harvard University Press.

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

Knauff, M., Rauh, R., Schlieder, C., & Strube, G. (1998). Mental models in spatial reasoning. In C. Freksa, C. Habel, & K. F. Wender (Eds.), *Spatial cognition: An interdisciplinary approach to representation and processing of spatial knowledge* (pp. 267–291). Berlin, Germany: Springer.

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review, 120*, 561-588.

Rauh, R., Hagen, C., Knauff, M., Kuß, T., Schlieder, C., & Strube, G. (2005). Preferred and alternative mental models in spatial reasoning. *Spatial Cognition and Computation, 5,* 239–269.

Schaeken, W., Van der Henst, J.-B., & Schroyens, W. (2007). The mental models theory of relational reasoning: Premises' relevance, conclusions' phrasing, and cognitive economy. In W. Schaeken, A. Vandierendonck, W. Schroyens, & G. d'Ydewalle (Eds.), *The mental models theory of reasoning: Refinements and extensions* (pp. 129–149). Mahwah, NJ: Erlbaum.

Vandierendonck, A., De Vooght, G., Desimpelaere, C., & Dierckx, V. (1999). Model construction and elaboration in spatial linear syllogisms. In W. S. Schaeken, G. De Vooght, A. Vandierendonck, & G. d'Ydewalle (Eds.), *Deductive reasoning and strategies* (pp. 191–207). Mahwah, NJ: Erlbaum.

Vandierendonck, A., Dierckx, V., & DeVooght, G. (2004). Mental model construction in linear reasoning: Evidence for the construction of initial annotated models. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 57*A, 1369–1391.