Similarities and Individual Differences in the Wason Selection Task: An Item Response Theory Analysis

Kuninori Nakamura (knaka@ky.hum.titech.ac.jp)
Japan Society for the Promotion of Science
Graduate School of Decision Science & Technology, Tokyo Institute of Technology
2-12-1, Ohkayama, Meguro-Ku, Tokyo 152-8552, Japan

Abstract

The four-card selection task (Wason, 1966) has been one of the most well-known tasks used in the literature on human reasoning. This article aimed to analyze this selection task by item response theory (Lord & Novick, 1968). Japanese undergraduates (N = 327 and 277 in Studies 1 and 2, respectively) responded to up to 10 types of representative Wason selection tasks, including the indicative task (Wason, 1966), beer task (Griggs & Cox, 1982), and cassava task (Cosmides & Tooby, 1989). The results of the analysis by the two-parameter logistic model indicated the following: the indicative task was similar to the beer and cassava tasks in terms of the discrimination parameter, and relative difficulty between the tasks would vary according to the value of theta, estimated by the two-parameter logistic model.

Keywords: Four-card selection task; item response theory; individual difference

Introduction

The Wason selection task (Wason, 1966) has been one of the most well-known tasks used in the literature on human reasoning. In its original version, the participants are presented with a conditional rule of the form, "if P, then Q," and four cards with information about P on one side and information about Q on the other side. The visible sides of the four cards display the information, P, not-P, Q, and not-Q. The participant's task is to indicate which of the four cards needs to be turned over in order to determine whether the rule has been violated. Although the Wason selection task is very simple, it is well-known because of the low percentage of the correct response to the original version of this task (Wason, 1966), or the thematic content effect (Wason & Shapiro, 1971; Johnson-Laird, Legrenzi, & Legrenzi, 1972).

The purpose of this study is to explore the latent factor behind the Wason selection task. In doing so, we briefly review the literature on selection tasks and point out that a quantitative multivariate analysis fruitfully contributes to this literature. Subsequently, we argue that item response theory (IRT: Lord & Novick, 1968) is adequate to describe the latent factor behind the various types of selection tasks. We also report two studies that analyzed selection tasks by IRT and concluded that the selection tasks are similar to each other, and individual differences in logical reasoning ability may affect participants' interpretations of the thematic content effect (Wason & Shapiro, 1971; Johnson-Laird, Legrenzi, & Legrenzi, 1972). This provides important theoretical suggestions to the domain of human reasoning.

The Wason selection task and its variations

It is well known that people often deviate from the normative principle when solving logical reasoning tasks. The most impressive example of the irrationality of human reasoning in the logical domain is the Wason selection task (Wason, 1966). In the original version of this task, participants are shown four cards, two of which display a letter and the other two, a digit. They are told that all four cards have a letter on one side and a digit on the other. Further, they are given a certain rule such as "if a card has a vowel on one side, it has an even number on the reverse." The participants are then asked to turn over those cards that they believe will determine whether the rule is true. To test the rule appropriately, participants must check "E" and "7." Despite the apparent simplicity of the selection task, it is notoriously difficult for people to solve correctly. Typically, less than 10% of the participants are able to determine the logically correct solution to the task.

It is also known that a participant's performance on the selection task increases substantially when the content of the conditionals is more concrete. For example, in Griggs and Cox's (1982) experiment, the participants were required to test the rule "if one drinks beer, he must be over 20," and were shown four cards that were assumed to represent a drink and an age as follows: "beer," "coke," "21," and "18." Griggs and Cox (1982) reported that the participants' performance improved substantially, and the same findings have been replicated by others (e.g., Cheng & Holyoak, 1985; Cosmides, 1989). This phenomenon is called the thematic content effect (also see Wason & Shapiro, 1971; Johnson-Laird, Legrenzi, & Legrenzi, 1972). To account for the thematic content effect, researchers have proposed various theories such as pragmatic schema theory (Cheng & Holyoak, 1985) and social contract theory (Cosmides, 1989).

Recently, it has become common to classify selection tasks into two categories: indicative tasks and deontic tasks (e.g., Evans & Over, 1996). According to this classification, indicative tasks primarily reflect the logical aspect of reasoning while deontic tasks require reasoning on whether a social rule is satisfied. Although no consensus has been reached with respect to the content of the deontic rule (e.g., Cheng & Holyoak, 1985; Cosmides, 1989; Fiddick, 2004), most researchers agree that there is a clear distinction between indicative and deontic tasks.

In addition, several researchers (Kirby, 1994; Manketlow & Over, 1991; Oaksford & Chater, 1994) have pointed out a possibility that the selection task is not a mere logical

reasoning task. Looking at its format, it is clear that the selection task is a logical reasoning task because the rule to be tested ("if P, then Q") itself is logical conditional. However, the selection task also requires participants to choose the cards to be turned in order to test the rule. Therefore, the selection task can be considered to have two aspects: interpretation of the conditionals and decision making with regard to which cards to select. Some previous studies (e. g., Kirby, 1994) have demonstrated that the decision making aspect of the selection task affects the performance of participants by manipulating the probabilistic aspect of the content of the conditional.

Two unexplored issues with regard to selection tasks: similarities and individual difference

Although many findings have been reported, there remain a number of unanswered questions selection tasks. One limitation of prior studies is that they rely on the comparison between tasks to explore the latent nature of selection tasks. When investigating selection tasks, the previous studies mainly employed two or more kinds of selection tasks, in which the content of the conditional written on the cards was manipulated, and compared the performance on the tasks, including the percentage of the correct response. For example, to explore social contract theory (Cosmides, 1989), which predicts the elicitation of the correct response in selection tasks by the conditionals containing some kind of social contract ("if you avail of the benefits, you must pay the obligation."), researchers have prepared two or more types of selection tasks, manipulated with regards to whether the content of the conditionals contain social contract, and then compared performance on the tasks, such as the percentage of the correct response (Griggs & Cox, 1982; Kirby, 1994) or the pattern of the incorrect responses (e. g., Oberauer, Wilhelm, & Diaz, 1999). The comparisons between tasks have long been a common method in the research domain of selection tasks.

However, two research issues remain unexplored. First, the comparison between tasks cannot treat individual differences in the response to the selection task. Although previous studies have reported many findings such as the thematic content effect (Wason & Shapiro, 1971; Johnson-Laird, Legrenzi, & Legrenzi, 1972) or matching bias (Evans & Lynch, 1973), it is difficult to consider these findings as relevant to the whole population. Rather, it is natural to assume individual differences in participants' interpretations or answers to the selection tasks. For example, there is a possibility that some tasks are relatively easy for some participants and not quite so for others. In this situation, a comparison between tasks is not adequate to describe the nature of the tasks.

Second, although a comparison can reveal the differences between tasks, it is not adequate to uncover the similarities among them. For example, all the selection tasks consistently require participants to interpret the meaning of the conditionals "if P, then Q." Accordingly, selection tasks can be considered to reflect logical reasoning ability. If this is the case, how are they similar to each other? Rather, how do they differ in reflecting logical ability? The studies in the past have not paid attention to this matter.

Stanovich and West's (1998a, b, 2000) studies are the few exceptions in this regard. Stanovich and West (1998a) explored the relationships among the various kinds of selection tasks and reported significant correlations. Stanovich and West (1998b) investigated the relationship between the Scholastic Achievement Test (SAT) and the performance in selection tasks, and found significant differences between the SAT scores of participants who could solve the task and those who could not. These studies suggest that the various kinds of selection tasks were similar to each other, and the similarities among them can be explained by an ability that the SAT measures. However, Stanovich and West's studies were limited to correlations between the specific combinations of selection tasks.

Thus, although many studies have been conducted to explore the cognitive processes behind the selection tasks, they did not investigate the similarities between them. In addition, they did not treat the problem of individual differences, which would affect the interpretation of phenomena concerning selection tasks, such as the thematic content effect. This study investigates these problems by IRT, which enables the examination of individual difference in participants' responses to tasks. We now show how IRT treats the issues of similarities and individual difference.

Item response theory

IRT is a measurement framework that models how individuals respond to individual items, and can provide research tools that are able to answer specific questions regarding the latent trait measure, through IRT affect responses to the item. In IRT, individuals are described by a latent construct, commonly denoted as theta (θ), which determines the probability of affirming items within a scale. θ is typically assumed to be normally distributed, $N(0,1)$. Items are described by item response functions (IRFs), which are trace lines that relate particular values of θ to a specified probability of affirming the item (see Fig. 1 for examples of IRFs). The IRFs are determined by estimated parameters from a model that is chosen by the researcher. A common model for dichotomous items is the two-parameter logistic (2PL) model, described as

$$P(u_i = 1 \mid \theta) = \frac{1}{1 + \exp(-1.7a_i(\theta - b_i))}, \qquad (1)$$

where the probability that a person with a latent trait affirms an item is a function of two parameters: ai, a discrimination parameter, and bi, a threshold parameter. The discrimination parameter represents how the latent trait determines the response choice. The threshold parameter expresses the difficulty level of an item indicating the person's theta value, whose probability to affirm the item is 0.5.

By using the 2PL model, we can investigate various aspects of selection tasks that could not be treated previously. Specifically, this model can represent that the

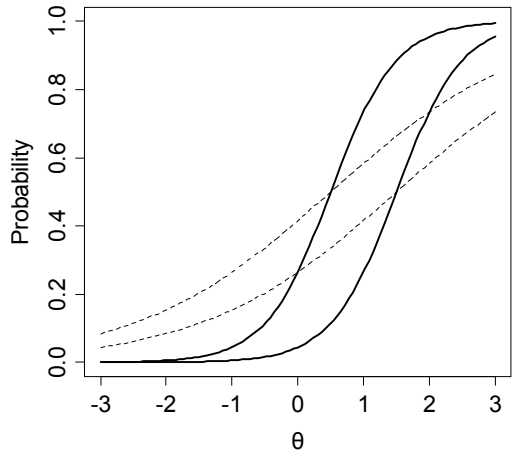

Figure 1 Examples of item response functions (IRF): Bold lines show IRFs with high discriminative parameters (1.2) and dotted lines show IRFs with low discriminative parameters (0.4).

difficulty level of the items would vary according to individual differences in participants' theta values. Figure 1 shows four response curves with different values of discrimination and threshold parameters. With regard to these four items, a person with a low theta value will easily answer the item with a low value of the discrimination parameter and vice versa for a person with a high theta value. This relationship indicates that the relative difficulty of an item would change according to the theta value of the person.

This ability of the 2PL model to represent individual difference is important when interpreting the difference in performance on selection tasks. As stated above, previous studies have employed the method of comparison among tasks, paying special attention to the difference in the percentages of the correct response. However, the IRFs in Figure 1 suggest that the relation between two tasks would vary according to θ. In other words, the 2PL model can express that the thematic content effect depends on individual differences between the participants. In this regard, the 2PL model is adequate to investigate the relationship between the thematic content effect and the individual differences between participants.

The 2PL model also enables a discussion on the similarities among tasks through an examination of the discriminative parameter. The discriminative parameter shows how the latent trait determines the response choice. In other words, it represents how the items reflect the latent trait. Thus, we can argue the similarities among selection tasks by investigating how they relate to logical reasoning ability. For example, with regards to the four IRFs in Figure 1, we can say that the bold lines are similar to each other and the dotted ones are also similar to each other in how they reflect the latent trait.

Table 1 The selection tasks used in Studies 1 and 2

						Correct response (%)		Factor load	
		p	not-p	q	not-q	Study1	Study2	Study1	Study2
1	Vowel	E	K	4	7	9.6	8.33	0.84	0.82
2	A	A	T	2	9	6.50	5.80	0.91	0.85
3	Beer	Beer	Coke	22	15	52.3	47.46	0.87	0.91
4	Envelope	Closed	Not closed	20$	10$	14.6	14.13	0.83	0.67
5	Cassava (Deontic)	Cassava	Moronut	Tatoo	No tatoo	54.2	41.30	0.86	0.79
6	Cassava (Descriptive)	Cassava	Moronut	Tatoo	No tatoo	11.1	12.32	0.56	0.68
7	Not-A	A	T	2	9		31.52		0.66
8	Cholera	Depart	Enter	Colera	(none)		24.64		0.74
9	Employee	Worked	Did not worked	Get a day off	Did not get a day off		25.00		0.66
10	Employeer	Worked	Did not worked	Get a day off	Did not get a day off		26.81		0.66

In sum, the above discussion shows that IRT can solve the issues related to the similarities and individual differences among selection tasks. Consequently, the purpose of this study is to analyze selection tasks by the 2PL model. Study 1 employed six types of selection tasks including indicative and deontic tasks. Study 2 aimed at replicating the findings of Study 1 using ten types of selection tasks.

Study 1

Method

Participants: Three hundred and twenty-seven undergraduates who were new to selection tasks participated in the study as part of a course credit.

Materials and procedure

We used a total of six types of selection tasks including the two types of indicative tasks (Wason, 1966), the beer task (Griggs & Cox, 1982), the post task (Wason & Shapiro, 1971), and the two types of cassava tasks (Cosmides, 1989). One of the two types of cassava tasks corresponds to deontic tasks that contains social contract context, whereas the other corresponds to descriptive tasks that has no deontic context. Table 1 presents the precise descriptions of each task.

All the materials and instructions were provided to the participants in a booklet. For each task, the participants were shown four cards in the booklet. They were instructed to indicate which of the four cards needs to be turned over in order to determine whether the rule has been violated. Participants responded by checking the four cards that illustrated the conditionals "P," "not P," "Q," and "not Q." All the participants completed the tasks within 20 minutes.

Results and discussion

Table 1 shows the percentages of the correct response to the selection tasks. The two indicative tasks elicited very low percentages of correct responses (9.6% and 6.5% in the "if vowel" and "if A" tasks, respectively) whereas the deontic tasks elicited higher percentages of correct responses (52.3% and 54.2% in the beer and cassava tasks, respectively). The percentage of correct participants in the post task was higher than those in the indicative tasks, although, its value was relatively low (14.6%). Thus, we can conclude that this study replicated the thematic content effect. In sum, Study 1 succeeded to replicate the previous findings with respect to selection tasks.

We then performed a factor analysis of the tetrachoric correlation matrix of the six tasks with varimax rotation to test the unidimensionality of the selection tasks before applying the 2PL model. Eigenvalues of the factor analysis were 4.18, 0.76, 0.60, 0.25, 0.20, and 0.02 for the first to the sixth factor solutions, respectively, indicating that the first factor accounted for most of the variance in the six tasks. Thus, we can conclude that the unidimentionality of the tasks was supported; the six selection tasks can be placed on a unidimensional continuum labeled *logical reasoning ability*, regardless of whether they are indicative or deontic. In addition, factor loading to all the tasks were high, indicating that performance on the tasks was strongly affected by logical reasoning ability, and that the selection tasks were very similar to each other in reflecting logical reasoning ability. This unidimensionality and similarity among the tasks themselves was very surprising, because previous studies pointed out a difference between the indicative and deontic tasks (e. g., Cosmides, 1989; Griggs & Cox, 1982).

On the basis of the results of the factor analysis, we applied the 2PL model to explore how individual difference affects performance on selection tasks. The IRFs shown in Figure 2 demonstrate that apart from the descriptive cassava task, the selection tasks used in Study 1 were very similar to each other in terms of the magnitude of the discriminative parameter. That is, although they differ in their threshold parameter values, performance was nearly the same, reflecting the participants' ability to solve logical reasoning tasks.

Specifically, similarities of the discrimination parameters between the indicative and deontic tasks were noteworthy. The IRFs of these tasks indicate that the thematic content effect would occur independent of the individual difference in reasoning ability because there was no crossover between indicative and deontic tasks. This result also suggests that the thematic content effect would occur robustly regardless of the logical reasoning ability.

One more important point is the shape of the IRF of the descriptive cassava task. Owing to the low value of its discriminative parameter, the rank order of the probability of correct responses between the descriptive cassava task and the indicative tasks changed according to the participants' ability. That is, when θ was relatively low, the content of the descriptive cassava task elicited the correct response whereas when θ was relatively high, it had the opposite effect. This suggests the possibility that differences between the tasks would vary according to the individual difference in θ.

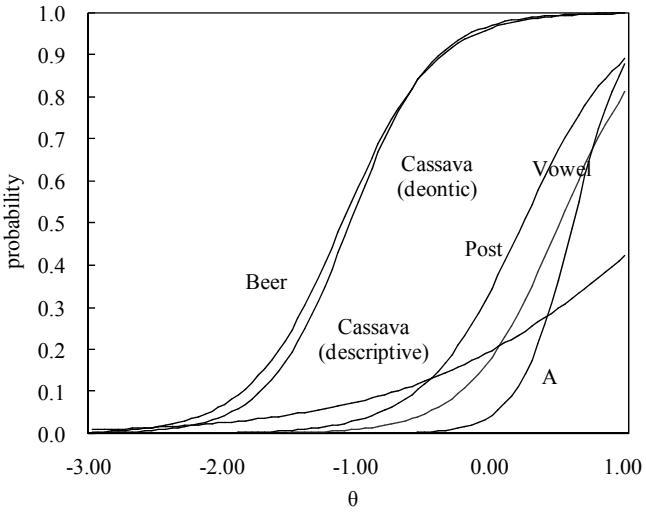

Figure 2 IRFs of the selection tasks used in Study 1

In sum, the findings in Study 1 indicate the following three points. First, the various kinds of selection tasks were similar to each other, regardless of the content of the conditional written on the cards. Second, the thematic content effect found in the previous studies would be a robust phenomenon, independent from the individual differences between participants. Finally, there is a possibility that some differences in the percentages of the correct response would be due to the responder's ability to solve the logical reasoning task, suggesting a danger with inferring the nature of tasks only from a comparison between them.

Study 2

One limitation of the first study 1 is its scarceness of the number of the tasks. Study 2 aimed to replicate the findings in Study 1 using more selection tasks. In Study 2, we used a total of ten selection tasks and examined whether the unidimensionality among them could be found.

Materials and procedure

Two hundred and seventy-seven undergraduates who were new to selection tasks participated in the study as part of a course credit. Table 1 shows the four new selection tasks employed in Study 2. In addition to the six tasks, the four new tasks were the negation (NA) task (Evans & Lynch, 1973), the cholera task (Cheng & Holyoak, 1985), and the two tasks used in Gigerenzer and Hug (1992). The procedure was almost the same as that followed in Study 1: all the materials and instructions were provided to the participants in a booklet, and participants were instructed to

indicate which of the four cards needs to be turned over in order to determine whether the rule has been violated.

Results and discussion

The percentages of the correct response are shown in Table 1. The two indicative tasks elicited very low percentages of correct responses (8.3% and 5.9% in the "if vowel" and "if A" tasks, respectively) whereas the deontic tasks elicited higher percentages of correct responses (47.5% and 41.3% in the beer and cassava tasks, respectively). Thus, we can conclude that this study replicated the thematic content effect. We also found the matching bias because the NA task elicited a higher percentage of correct responses (31.5%) than the indicative tasks. With regard to the two tasks employed in Gigerenzer and Hug (1992), although not a prominent difference, there was a slight increase in the number of correct responses that was predicted from the perspective change effect (Gigerenzer & Hug, 1992). Thus, we can conclude that Study 2 replicated the trends found in the previous studies.

Like in Study 1, we then performed factor analysis with varimax rotation to test the unidimensionality of the ten selection tasks. Eigenvalues of the factor analyses were 5.99, 1.36, 0.78, 0.66, 0.47, and 0.35 for the first to the sixth factor solutions, respectively, indicating that the first factor solution accounted for most of the variance in the tasks. Thus, the unidimensionality of the ten selection tasks was supported in Study 2. In addition, factor loading to the selection tasks were high (0.66–0.91), indicating that performance on the tasks was strongly affected by logical reasoning ability. Specifically, the factor loads to the two types of deontic tasks were very high (0.91 and 0.71 in the beer and cassava tasks, respectively). These results support the finding in Study 1 that selection tasks, whether indicative or deontic, can be aligned on the dimension of logical reasoning ability, and as a whole, they are similar to each other in the reflection of the logical reasoning ability.

We then applied the 2PL model to the ten selection tasks. The IRFs shown in Figure 3 demonstrate the same trends found in Study 1: the shapes of the IRFs of the indicative and deontic tasks, including the beer and deontic cassava tasks, were very similar to each other, while that of the descriptive cassava task was different. These results also support the robustness of the thematic content effect found in Study 1.

As a whole, the IRFs of the selection tasks other than the indicative and representative deontic tasks show similar Figure 3 IRFs of the selection tasks in Study 2 shapes. The discriminative parameter values of these tasks were lower than the two indicative and deontic tasks, and there were intersects of the IRFs of these tasks with those of the indicative tasks. Thus, Study 2 also replicated the finding of Study 1 that differences in the correct responses between the tasks would vary according to the responder's ability to solve logical reasoning tasks.

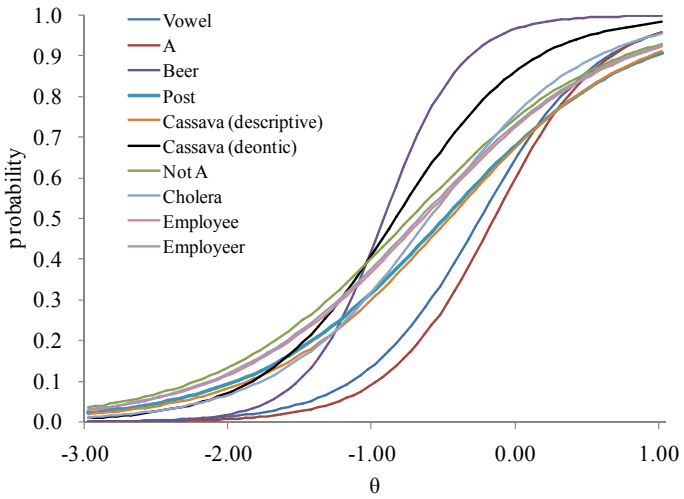

Figure 3 IRFs of the selection tasks in Study 2

Conclusion

Implications of the present study can be summarized as follows. The first implication entails that selection tasks mainly reflect the logical reasoning ability to interpret the meaning of the conditionals and the content written on the cards. This implication is slightly counterintuitive because recent studies have emphasized that selection tasks appear to reflect many aspects. As a result of the findings of theses studies, the selection task has come to be acknowledged as a task that enables us to explore various cognitive processes such as deontic reasoning (Cheng & Holyoak, 1985; Cosmides, 1989), decision making (Manketlow & Over, 1991), or dual processes (Evans & Over, 1996). Specifically, because of the impact of social contract theory (e. g., Cosmides, 1989) or optimal data selection theory (Oaksford & Chater, 1994), the logical aspect of selection tasks appears to have been underestimated. However, this study demonstrates that when estimated quantitatively, selection tasks are, primarily, logical reasoning tasks, regardless of the content of the conditionals. Specifically, the similarity between the indicative and deontic tasks in the reflection of logical reasoning ability is important because deontic tasks mainly require the estimation of the utility of a card rather than logical reasoning (Manketlow & Over, 1991).

The second implication involves the necessity to consider individual difference in logical reasoning ability to interpret the differences between tasks. This finding is important because recent studies on human reasoning employ selection tasks for a variety of purposes such as inter-cultural comparisons (Sugiyama, Tooby, & Cosmides, 2002), specifying the meaning of "deontic" (Fiddick, 2004), or neuroimaging analyses (e. g., Stone, Cosmides, Tooby, Kroll, & Knight, 2002). These studies used newly composed versions of selection tasks and employed the method of comparison between tasks to address their research issue. However, as the results of the 2PL models indicate, the

differences in the correct response between tasks would vary according to the logical reasoning ability.

Although this implication appears to demarcate a boundary condition of the thematic content effect, this study also supports the robustness of the thematic content effect found in previous studies. As stated above, the thematic content effect refers to an increase in performance when the deontic conditionals are employed. As the IRFs in Figures 2 and 3 show, the ordinal relationship between indicative tasks and deontic tasks was constant, independent of logical reasoning ability. That is, the deontic tasks were easier than the indicative tasks, regardless of the participant's logical reasoning ability. Thus, the implication on individual difference also supports the generality of the thematic content effect or the universality of social contract theory.

In sum, the application of the 2PL model to selection tasks enables us to explore various unresolved aspects. The 2PL model may shed light on the latent trait concerning issues like the similarity among tasks or individual differences in not only selection tasks but also other reasoning tasks such as probability judgment (e. g., Tversky & Kahneman, 1973) or preference choice (e. g., Kahneman & Tversky, 1979). In fact, Zicker and Highhouse (1998) used the 2PL model to investigate individual difference in the framing effect, and showed that a latent construct labeled preference for risk was influential in predicting risky choice. We believe that the application of IRT to the domain of human reasoning will fruitfully contribute to the determination of factors that affect how people think.

Finally, future research should be devoted to developing the construct validity of logical reasoning ability. The analysis in this study focused on the internal analysis of items and did not consider relations with other scales and behaviors. Therefore, a limitation of this research is the arbitrariness in labeling the underlying latent trait measured by selection tasks. Thus, further construct validation research needs to be conducted to develop a precise label for this dimension.

Acknowledgment

This research was supported by a grant from the Japan Society for the Promotion of Science.

References

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*, 391-416.

Cosmides, L. (1989). The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*, 187-276.

Evans, J. St. B. T., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology*, **64**, 391-397.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.

Fiddick, L. (2004). Domains of deontic reasoning: resolving the discrepancy between the cognitive and moral reasoning literatures. *Quarterly Journal of Experimental Psychology*, *57(A)*, 447-474.

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: social contracts, cheating, and perspective change. *Cognition 43*,127-171.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, *73*, 407-420.

Johnson-Laird, P. N., Legrenzi,P., & Legrenzi, S. M. (1972). Reasoning and a sense of reality. *British Journal of Psychology,* **63**, 395-400.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica, 47*, 263-91.

Kirby, K. N. (1994). Probabilities and utilities of fictional outcomes in Wason's four-card selection task. *Cognition*, *39*, 85-105.

Lord, F. N., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Manketlow, K. I., & Over, D. E. (1991). Social roles and utilities in reasoning with deontic conditionals. *Cognition*, *39*, 85-105.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*, 608-631.

Oberauer, K., Wilhelm, O., & Diaz, R. R. (1999). Bayesian rationality for the Wason selection task? A test of optimal data selection theory. *Thinking and Reasoning*, *5*, 115-144.

Stanovich, K. E., & West, R. F. (1998a). Cognitive ability and variation in selection task performance. *Thinking and Reasoning, 4*, 193-230.

Stanovich, K. E., & West, R. F. (1998b). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*, 161-88.

Stanovich, K. E., & West, R. F. (2000). Individual difference in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*, 645-726.

Stone, V., Cosmides, L., Tooby, J., Kroll, N., & Knight, R. (2002). Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *Proceedings of the National Academy of Sciences, 99*, 11531-11536.

Sugiyama, L., Tooby, J., & Cosmides, L. (2002). Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. *Proceedings of the National Academy of Sciences*, *99*, 11537-11542.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 4*, 207-232.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology*. Harmondsworth: Penguin.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, *23*, 63-71.

Zicker, M. J., & Highhouse, S. (1998). Looking closer at the effects of framing on risky choice: an item response theory analysis. *Organizational Behavior and Human Decision Processes*, *75*, 75-91.