岩手大学教育学部研究年報　第 71 巻 (2012. 3) 33 ～ 43

Hypothesis Testing Strategy in Wason’s 2-4-6 Reasoning Task

Nobuyoshi IWAKI＊・Megumi TSUDA＊＊・Rumi FUJIYAMA＊＊・Ryo AKASAKA＊＊＊

(Received Jan. 12, 2012)

Abstract

We examined theories of hypothesis testing strategy in Wason’s 2-4-6 task. Twenty-one undergraduate students were given single trials that featured a hypothesis and an example conforming to the hypothesis, and they received “yes/ no” feedback as to whether the example conformed to a hidden rule. We included three conditions: mere confirmation, inconsistency, and consistency. “Yes” feedback was presented in the former two conditions. Feedback negativity, a negative-event-sensitive event-related potential, was enhanced in inconsistency trials that featured falsification of contrary hypotheses, in comparison with mere confirmation trials. This result is consistent with the “iterative counterfactual strategy” (Oaksford & Chater, 1994) account but not with the “positive test heuristic” (Klayman & Ha, 1987), in that the latter is unable to qualitatively discriminate “yes” feedback trials.

Keywords: Wason’s 2-4-6 task, hypothesis testing strategy, iterative counterfactual strategy, positive test heuristic, event-related brain potential

Correcting erroneous thoughts in a reasoning task is an important aspect of the adaptive nature of human cognition. This sort of correction process can be examined using Wason’s 2-4-6 task (Wason, 1960).

Wason’s 2-4-6 task requires participants to figure out a hidden rule about number trios that the experimenter has in mind (e.g., ascending numbers). The participants are initially given the example “2, 4, 6” , which can be deduced from a rule, and are told that the example conforms to a specific rule. A participant might then hypothesize the rule to be “even numbers increasing by two” , and then tests the hypothesis by presenting an example such as “6, 8, 10” to the experimenter. Participants then receive “yes/ no” feedback regarding whether the example in fact conforms to the rule. Given that the rule is “ascending numbers” , “yes” feedback is given when the participant’s example is “6, 8, 10” . This sequence is defined as a single trial. If participants are sure of the hypothesis, they report it to

＊　Nobuyoshi　IWAKI　Faculty of Education, Iwate University
＊＊　Megumi TSUDA・Rumi FUJIYAMA　Faculty of Humanities, Kyushu Lutheran College
＊＊＊　Ryo AKASAKA　Faculty of Engineering, Kyushu Sangyo University

the experimenter directly and are given “yes/ no” feedback regarding whether it is correct. When the feedback is “no” , they are required to continue on with the task sequence. The 2-4-6 task therefore enables researchers to capture the thinking process by which participants test their own ideas and revise them accordingly.

Previous studies have shown that participants are likely to test a given hypothesis with a positive example that can be deduced from the hypothesis (for reviews, see Evans, 1989; Gorman, 1995). For example, when the hypothesis is “even numbers increasing by two” , participants are likely to present a positive example (such as “6, 8, 10”) conforming to the hypothesis. While this typical response seems to originate from a behavioral tendency to seek information that confirms one’s own beliefs (e.g., Wason, 1960), we could instead consider this response as the result of using a disconfirmation strategy.

There are cases where participants aim to provide a falsification by demonstrating that two contrary hypotheses can be confirmed. This is a counterfactual reasoning strategy that is logically the same as “reductio ad absurdum” rather than “modus tollens” . With regard to the 2-4-6 task, Oaksford and Chater (1994) proposed an “iterative counterfactual strategy” (ICS), a proposal that originated with Farris and Revlin (1989a, 1989b). According to this strategy, participants pay attention to number trios that conform to a true rule, and then provide a hypothesis that incorporates a property (or properties) common to the trios. For example, they may develop the idea of “even numbers that increase by two” after seeing “2, 4, 6” and “8, 10, 12” . Next, if a property such as “even numbers” is chosen, participants will seek to confirm its complement (or opposite, see Oaksford & Chater, 1994, footnote 7, p. 154), in this case by using odd number trios (such as “1, 3, 5”). If the feedback in this case is “Yes” , they can be sure that the issue of “even vs. odd” does not in fact relate to the true rule. Caverni & Rossi (1997) reported that participants frequently tested contrary hypotheses, such as even versus odd numbers, findings that seem to be compatible with the ICS theory (although they did not note this).

There is at least one other theoretical possibility. Klayman and Ha (1987) pointed out that falsification is enabled by positive feedback, if considering possible denotation relationships between the sets of number trios that can be categorized according to a true rule or hypothesis. For example, given that the rule is “even numbers” , a hypothesis of “numbers increasing by two” includes the set of “even numbers increasing by two” , and therefore the denotation of the rule may partially overlap with one of the hypotheses. One type of falsification is a case where a negative example (e.g., 6, 4, 2) of a hypothesis (numbers decreasing by two) also serves as a positive example of the “even numbers” rule, while in another case a positive example (e.g., 1, 3, 5) of the hypothesis also serves as a negative example of the rule. In sum, according to Klayman & Ha (1987, 1989), falsification is realized not only by negative but also by positive examples; in our daily lives, positive examples are more likely to provide falsification than are negative examples (Klayman & Ha, 1987). The notion of a “positive test heuristic” helps us to understand our tendency to pay attention to positive examples while also accounting for how we are able to disconfirm hypotheses.

The ICS and positive test heuristic accounts provide different considerations of trials that feature a positive example. In terms of the ICS, positive examples that confirm an alternative hypothesis (the complement/opposite of an original hypothesis) lead to disconfirmation of both hypotheses according to the logic of ‘reductio ad absurdum’ . Although the confirmation of an original hypothesis is mere confirmation, confirmation of the alternative hypothesis falsifies both contrary hypotheses. On the other hand, the positive test heuristic account is unable to differentiate these types of confirmation. The aim of the present study was to investigate which theory is persuasive.

We propose that the qualitative difference in trials featuring a positive example may be reflected in brain activity, and that the event-related brain potential (ERP) dubbed feedback negativity (FN) provides a useful measure of activity reflecting feedback recognition (a hypothesis is evaluated using feedback information provided by the experimenter).

In a time-estimation task, Miltner, Braun, and Coles (1997) observed that negative feedback that denoted a time-estimation error elicited the FN, and its amplitude was influenced by neither sensory nor response modalities. The FN appeared predominantly at fronto-central scalp sites, and reached a maximum of around 250 ms after feedback presentation. The FN is thought to originate from the anterior cingulate cortex (see also Ruchsow, Grothe, Spitzer, & Kiefer, 2002). In a gambling task involving rewards and losses, the FN increased when negative feedback denoting a loss was provided (Gehring & Willoughby, 2002). Its amplitude was context-dependent in that it was sensitive to relative losses rather than the absolute magnitude of the loss (Holroyd, Larsen, & Cohen, 2004), but FN is not influenced by stimulus frequency (Hajcak, Holroyd, Moser, & Simons, 2005). In addition, FN enhancement has been observed in a card game in which participants guess which card will appear out of four possibilities (Ruchsow et al., 2002), as well as in a modified Wason's 2-4-6 task (Papo, Baudonnière, Hugueville, & Caverni, 2003). Papo's et al. (2003) study, however, could not differentiate the positive feedback trials because the researchers did not control participants' generation of hypotheses. In any case, FN amplitude enables us to measure the differential brain activities elicited by feedback.

According to the positive test heuristic account (Klayman & Ha, 1987), positive feedback trials that feature a positive example are all identical because feedback in such trials only indicates confirmation. This account does not predict any FN amplitude differences for such trials. However, according to the ICS account, in those cases where positive feedback leads to disconfirmation through counterfactual reasoning, brain activity should be distinct from that observed during mere confirmation trials. As positive feedback trials with counterfactual falsification (inconsistency trials) are supposed to have a negative value in that two contrary hypotheses should be eliminated, the FN should increase during these trials. However, if negative feedback is given during a counterfactual reasoning trial (or consistency trial), the FN amplitude, if any, will be the lowest among the trial conditions, given that a positive value accompanies the finding of true rule-related information. Finally, in mere confirmation trials, positive feedback does not help participants to judge whether the numerical property under consideration relates to a true rule. The FN amplitude here, if any, should fall between those observed during the other conditions.

We conducted a modified Wason's 2-4-6 task. While participants usually generate hypotheses by themselves and report number trios, in order to better control the thinking process and measure electrical brain activity, we prepared both typical confirmation trials and trials in which the participants performed counterfactual reasoning.

Method

Participants

Twenty-one undergraduate students (20 women, 18-22 years old) voluntarily participated in this experiment. Informed consent was obtained from the participants before each testing session.

Task and procedure

Participants were each given a record sheet as shown in Figure 1, albeit without the explanation in parentheses. Initially, the property that participants needed to attend to was entered, which is common to “yes” feedback number trios. For example, in the case of “2, 4, 6” are all even” in the first row, a property (even) is focused upon and then the hypothesis of “even numbers” is entered. Because this is a mere confirmation trial, “Let’s confirm whether the numbers are even” was mentioned and followed by a positive instance: “8, 10, 12” and “yes” feedback. As the next row depicted is a trial in which counterfactual reasoning is applied, “Then, how about odd numbers?” was entered, followed by a positive example: “1, 3, 5” .

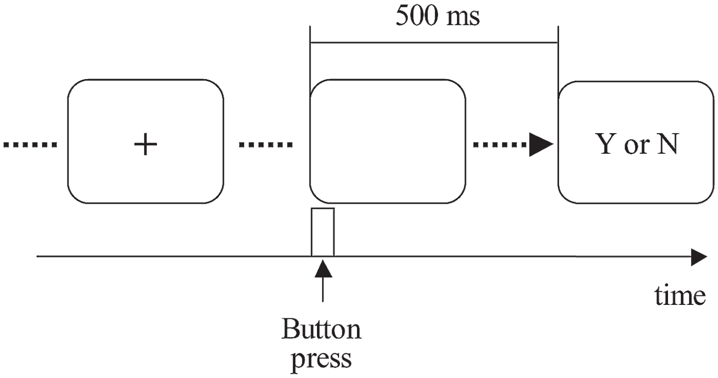

Figure 2 Feedback presentation procedure

In order to control participants’ use of strategies, they were required to read aloud each row before receiving feedback, as if they were actually thinking the content. The feedback (Y/ N) regarding whether the number trios conformed to the rule was presented at the center of a display 50 cm in front of the participant. The feedback presentation sequence is shown in Figure 2. The feedback was provided for 1000 ms, 500 ms after the participant’s voluntary button press, and the participant checked “Yes/ No” in the column “Feedback” . At the end of each row, we required participants to judge whether the numerical property related to the true rule, in order to assess their understanding of the task. Each block consisted of 10 trials. The participants guessed at and wrote down the most promising hypothesis and an example conforming to the hypothesis in the 10th trial. For the sake of reconfirmation, participants were required

No.	What you found out	Hypothesis	What you will test	Example	Feedback	Relatedness to the true law
—		—	**(When the rule is 'positive numbers')**	2, 4, 6	Yes	
1	'2, 4, 6' are all even.	even numbers	Let's confirm 'even numbers' **(confirmation)**	8, 10, 12	Yes No	Yes No I don't know yet.
	All were even numbers heretofore.	even numbers	Then, how about 'odd numbers'? **(inconsistency)**	1, 3, 5	Yes No	Yes No I don't know yet.
	Numbers were positive heretofore.	positive numbers	Then, how about 'negative numbers'? **(consistency)**	-3, -2, -1	Yes No	Yes No I don't know yet.
10	Finally, confirm the hypothesis in mind.		Let's confirm this hypothesis. **(confirmation)**		Yes No	Yes No I don't know yet.

The final hypothesis in my mind : ______________________________ ⇒ Yes・No

Figure 1 An example of a record sheet

to write down their final hypothesis and were then given the feedback. The participants completed 20 blocks, after an initial 10 practice trials. Rules are shown in Table 1. Twenty rules were prepared and the items involved seven elementary rules (e.g., ascending numbers) and 13 elements-combined rules (e.g., positive and ascending numbers), each of which was adopted once per experimental session. The percentages for each trial condition were 28% for mere confirmation, 54% for inconsistency, and 18% for consistency.

Table 1 Rules. Twenty rules were prepared, including seven elementary rules and 13 elements-combined rules.

Any numbers
Ascending numbers
Constant interval
Positive numbers
Without decimals
Without fractions
Without ‘0’
Combinations of some properties mentioned above

Electrophysiological recording and data analysis

Electroencephalograms (EEGs) were recorded with Ag/AgCl electrodes from Fz, Cz, Pz, and the left ear lobe, with each being referred to the right ear lobe according to the 10-20 system (Jasper, 1958). Other electrodes were placed above and below the right eye and on the outer canthi of both eyes. A common electrode was placed on the forehead. The analogue signals were amplified (AB-610J, NIHON KODEN) and digitized on-line at 200 Hz. The electrode impedance was kept below 5 kΩ, and the bandpass frequency was 0.05-30 Hz.

The EEG signals were recalculated, each being referred to the algebraic average of the left and right ear lobe electrodes. The EEG and EOG (electrooculogram) signals during each 1000 ms were averaged starting at 200 ms (the baseline) before feedback. Trials with EEG or EOG exceeding ±50 μV were discarded before averaging.

Repeated measures analysis of variance (ANOVA) was used to analyze the data. The Greenhouse-Geisser correction for repeated measures was applied where appropriate. Tukey’s HSD test was adopted as a multiple comparison test ($p < .05$).

Results

Although one participant provided an incorrect answer during one block, all participants were considered to have performed the task correctly.

The number of trials averaged was 47 ± 7 (mean ± *SD*, range = 31-56) for the mere confirmation condition, 93 ± 11 (range = 70-108) for inconsistency, and 31 ± 3 (range = 25-35) for consistency.

Grand average waveforms are depicted in Figure 3. Negative potentials peaking at about 300 ms were observed predominantly at Fz and Cz, in particular during inconsistency trials. We calculated the average amplitudes between 250 ms and 350 ms for each participant (see also Table 2). A 3 (electrode location) × 3 (trial condition) ANOVA conducted on the mean amplitudes revealed significant main effects of electrode location (F (2, 40) = 7.63, $p < .01$, $\varepsilon = .67$, Fz < Cz; Fz < Pz) and trial condition (F (2, 40) = 11.61, $p < .01$, $\varepsilon = .62$, inconsistency < mere confirmation < consistency), as well as a significant interaction, F (4, 80) = 4.20, $p < .05$, $\varepsilon = .58$. A test of the simple main effect showed that at Fz and Cz the mean amplitudes were significantly different between each pair, but at Pz, the amplitudes were significantly different between inconsistency and the other two conditions.

Table 2 *Mean amplitudes (±SE) as a function of trial condition and recording site*

	Confirmation	Inconsistency	Consistency
FN			
Fz	−0.97±0.95	−3.47±1.08	1.76±1.22
Cz	0.99±0.84	−1.68±0.78	3.16±1.21
Pz	2.24±0.93	0.08±0.81	3.75±1.42
P3			
Pz	2.11±0.90	1.00±0.83	7.49±1.41

In addition, P3 was clearly observed for the consistency condition, so we calculated the average amplitudes between 350 ms and 450 ms at Pz and performed a one-way ANOVA on trial condition. This ANOVA revealed significant effects (F (2, 40) = 19.06, $p < .001$, $\varepsilon = .58$, mere confirmation < consistency, inconsistency < consistency).

Discussion

The purpose of the present study was to examine whether the positive test heuristic or the ICS better explains the tendency to use positive examples for hypothesis testing. From the standpoint of the positive test heuristic, "yes" feedback for a positive example indicates only "confirmation" , such that trials featuring "yes" feedback are all identical in quality. If this were the case, the FN amplitudes should not have been different across the mere confirmation and inconsistency conditions. However, as seen in Figure 3, grand average ERPs for mere confirmation and inconsistency trials appear to be different. A negative potential considered to be the FN based on polarity, latency, and topography clearly appeared during inconsistency trials and was significantly greater between 250 ms and 350 ms, in comparison

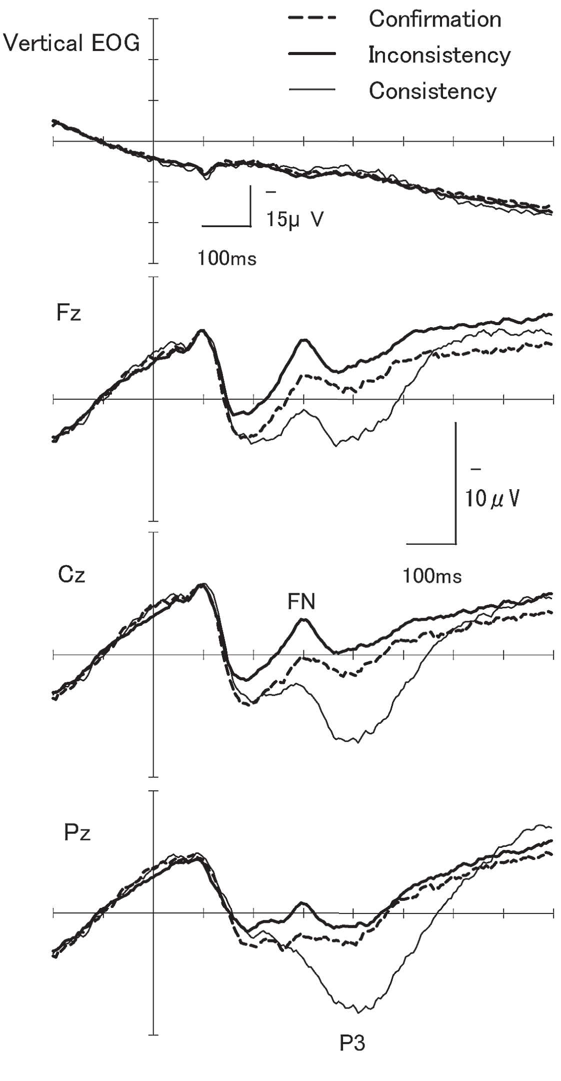

Figure 3　Grand average waveforms for each trial condition. The vertical line indicates the point in time of feedback presentation.

with mere confirmation trials. This result is not consistent with the positive test heuristic theory. We can therefore conclude that the positive test heuristic is not the best account of the tendency to use positive examples.

On the other hand, ICS theory helps us to better understand this result. "Yes" feedback leading to a disconfirmation through counterfactual reasoning allows participants to recognize the event as negative, given that both contrary hypotheses can be denied. The "yes" feedback in mere confirmation trials, however, does not negate a present hypothesis. Therefore, according to ICS theory, FN should be more robust in the inconsistency condition. Moreover, the ICS can explain the difference in amplitude between the mere confirmation and consistency conditions. In mere confirmation trials, strictly speaking, participants could not judge whether a numerical property under consideration pertained to a true rule, while during consistency trials, counterfactual reasoning enabled participants to consider "no" feedback as good/ positive, given that the finding of true rule-related information should be accompanied by positive value and/or emotion. It is therefore likely that the FN amplitude for the consistency condition should be smaller than that for the mere confirmation condition. However, it is unlikely that the ratio/ frequency for each trial condition influenced the differences in FN amplitude that we observed. The amplitude of ERPs is likely to increase when confronted with lower frequency information (for a review see Rugg & Coles, 1995), and moreover, Hajcak et al. (2005) found no influence of stimulus frequency on FN amplitude.

Hypothesis revision through the ICS is not revision through a mere mistake, but rather a hypothesis-construction based upon information that has survived and has been collected after testing hypotheses (this idea is similar to "model-building" as discussed in Klayman & Ha, 1989). The ICS is a strategy for finding task-related or -unrelated information through contingency judgments ("reductio ad absurdum") of contrary hypotheses and "yes/ no" feedbacks. This sort of strategy by which participants actively obtain information is not present in the positive test heuristic, which does not account for the notion of contingency judgments.

P3 is sensitive to stimulus frequency (Duncan-Johnson & Donchin, 1977). Although the P3 clearly appeared in the consistency condition, we could not explain this result from the viewpoint of the capturing of task-related information (true rule-related information in this study) (Johnson & Donchin, 1978; Gratton, Bosco, Kramer, Coles, Wickens, & Donchin, 1990). This is simply because the frequency of consistency trials was lower in the present study. In any case, the issue of event frequency does not seem to bear upon our main conclusions regarding the FN.

We can conclude that the ICS theory is consistent not only with behavioral evidence (Caverni & Rossi, 1997) but with ERP evidence as well, and can assume that participants are conscious of a contrary hypothesis and/or pay attention to a complement/opposite set. This mental process seems rational because (logically speaking) it is considered to be the application of "reductio ad absurdum" to find task-related or -unrelated information, through the confirmation of data deduced from contrary hypotheses. The following finding suggests that this mental process can be promoted: Feedback of DAX/ MED categories, as opposed to Yes/ No feedback, enables participants to perform better (Gorman, Stafford, & Gorman, 1987; Tweney, Doherty, Worner, Pliske, Mynatt, Gross, & Arkkelin, 1980; Vallée-Tourangeau, Austin, & Rankin, 1995; Wharton, Cheng, & Wickens, 1993). If the DAX/ MED categories promoted a bias toward paying attention to the complement/ opposite, it can be suggested that the mental process operating in such an experiment is basically identical to that operating during a standard 2-4-6 task (see also Vallée-Tourangeau et al., 1995). This identity issue regarding the mental processes operating across studies has yet to be examined.

References

Caverni, J-P. & Rossi, S. (1997). A nice bit of scandal: About a disconfirmation bias in the Wason's 2-4-6 problem. *Swiss Journal of Psychology, 56*, 239-242.

Duncan-Johnson, C. C. & Donchin. E. (1977). On quantifying surprise: the variation of event-related potentials with subjective probability. *Psychophysiology, 14*, 456-467.

Evans, J. St. B. T. (1989). *Bias in human reasoning*. Hove and London (UK): Rulerence Erlbaum Associates.

Farris, H. H. & Revlin, R. (1989a). Sensible reasoning in two tasks: Rule discovery and hypothesis evaluation. *Memory and Cognition, 17, 221-232*.

Farris, H. H. & Revlin, R. (1989b). The discovery process: A counterfactual strategy. *Social Studies of Science, 19*, 497-513.

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science, 295*, 2279-2282.

Gorman, M. E. (1995). Hypothesis testing. In S. E. Newstead & J. St. B. T. Evans (Eds), *Perspectives on Thinking and Reasoning*. Hove (UK) & Hillsdale (USA): Rulerence Erlbaum.

Gorman, M. E., Stafford, A., & Gorman, M. E. (1987). Disconfirmation and dual hypotheses on a more difficult version of Wason's 2-4-6 task. *Quarterly Journal of Experimental Psychology, 39A*, 1-28.

Gratton, G., Bosco, C. M., Kramer, A. F., Coles, M. G. H., Wickens, C. D., & Donchin, E. (1990). Event-related brain potentials as indices of information extraction and response priming. *Electroencephalography and clinical Neurophysiology, 75*, 419-432.

Hajcak, G., Holroyd, C. B., Moser, J. S., & Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology, 42*, 161-170.

Holroyd, C. B., Larsen, J. T., & Cohen, J. D. (2004). Context dependence of the event-related brain potential associated with reward and punishment. *Psychophysiology, 41*, 245-253.

Jasper, H. (1958). The ten twenty electrode system of the International Federation. *Electroencephalography and clinical Neurophysiology, 10*, 371-375.

Johnson, R., Jr. & Donchin, E. (1978). On how P300 amplitude varies with the utility of the eliciting stimuli. *Electroencephalography and Clinical Neurophysiology, 44*, 424-437.

Klayman, J. & Ha, Y.-W. (1987). Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review, 94*, 211-228.

Klayman, J. & Ha, Y.-W. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 15*, 596-604.

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. *Journal of Cognitive Neuroscience, 9*, 788-798.

Oaksford, M. & Chater, N. (1994). Another look at eliminative and enumerative behaviour in a conceptual task. *European Journal of Cognitive Psychology, 6*, 149-169.

Papo, D., Baudonnière, P-M., Hugueville, L., & Caverni, J-P. (2003). Feedback in hypothesis testing: An ERP study. *Journal of Cognitive Neuroscience, 15*, 508-522.

Ruchsow, M., Grothe, J., Spitzer, M., & Kiefer, M. (2002). Human anterior cingulated cortex is activated by negative feedback: Evidence from event-related potentials in a guessing task. *Neuroscience Letters, 325*, 203-206.

Rugg & M. G. H. Coles (Eds.) (1995). *Electrophysiology of mind: Event-related brain potentials and cognition*. New York: Oxford University Press.

Tweney, R. D., Doherty, M. E., Worner, W. J., Pliske, D. B., Mynatt, C. R., Gross, K. A., & Arkkelin, D. L. (1980). Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology, 32*, 109-123.

Vallée-Tourangeau, F., Austin, N. G., & Rankin, S. (1995). Inducing rule in Wason's 2-4-6 task: A test of the information-quantity and goal-complementarity hypotheses. *Quarterly Journal of Experimental Psychology, 48A*, 895-914.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12*, 129-140.

Wharton, C. M., Cheng, P. W., & Wickens, T. D. (1993). Hypothesis-testing strategies: Why two goals are better than one. *Quarterly Journal of Experimental Psychology, 46A*, 743-758.

This research was supported by a grant from the Ministry of Education, Culture, Sports, Science and Technology of Japan, Grant-in-Aid for Young Scientists (B), 15730306.