THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2002, *55A* (4), 1241–1272

A quantitative model of optimal data selection in Wason's selection task

Masasi Hattori

Ritsumeikan University, Kyoto, Japan

The optimal data selection model proposed by Oaksford and Chater (1994) successfully formalized Wason's selection task (Wason, 1966). The model, however, involved some questionable assumptions and was also not sufficient as a model of the task because it could not provide quantitative predictions of the card selection frequencies. In this paper, the model was revised to provide quantitative fits to the data. The model can predict the selection frequencies of cards based on a selection tendency function (STF), or conversely, it enables the estimation of subjective probabilities from data. Past experimental data were first re-analysed based on the model. In Experiment 1, the superiority of the revised model was shown. However, when the relationship between antecedent and consequent was forced to deviate from the biconditional form, the model was not supported. In Experiment 2, it was shown that sufficient emphasis on probabilistic information can affect participants' performance. A detailed experimental method to sort participants by probabilistic strategies was introduced. Here, the model was supported by a subgroup of participants who used the probabilistic strategy. Finally, the results were discussed from the viewpoint of adaptive rationality.

The four-card selection task was devised by Wason (1966) more than 30 years ago. Since then innumerable experiments and many theories explaining its various results have been published. In the original version of the selection task, an experimenter presents participants with a rule, "If there is a vowel on one side (p), then there is an even number on the other side (q)", and four cards showing an "E" (p), a "K" (*not-p*, usually expressed as $\neg p$), a "4" (q), and a "7" ($\neg q$). Each card has a letter on one side and a number on the other side. Participants are asked which cards they would need to turn over in order to determine whether the rule is true or false. The logically correct response is "E and 7" (p and $\neg q$), which is rarely made.

Requests for reprints should be sent to Masasi Hattori, College of Letters, Ritsumeikan University, 56-1 Kitamachi, Toji-in, Kita-ku, Kyoto 603-8577, Japan. Email: hat@lt.ritsumei.ac.jp

I would like to thank Mike Oaksford for many helpful comments and Osamu Yoshimura for valuable technical advice on an earlier draft of this paper. I am also grateful to Ken Manktelow, Nick Chater, David Over, and some anonymous reviewers for their valuable comments.

This paper was presented in part at the 14th Annual Meeting of the Japanese Cognitive Science Society (1997), the 61st Annual Meeting of the Japanese Psychological Association (1997), and the 2nd International Conference on Cognitive Science (Hattori, 1999).

http://www.tandf.co.uk/journals/pp/02724987.html DOI:10.1080/02724980244000053

Until recently, it was not clearly understood that there are two aspects to this task: a logical aspect concerning the rule and a decision-making aspect concerning the selection of cards. This duality makes it difficult both for participants to solve the task and for researchers to analyse it. Kirby (1994) brought the latter aspect to light by analysing the task using the notion of subjective expected utility. Oaksford and Chater (1994) successfully formalized the duality of this task in terms of Bayesian hypothesis testing (e.g., Earman, 1992; Howson & Urbach, 1989/1993). This computational model of human reasoning in the selection task is based on the concept of *rational analysis* (Anderson, 1990) and is called the *optimal data selection* (ODS) model. The model drew many criticisms (Almor & Sloman, 1996; Evans & Over, 1996a, b; Green & Over, 1998; Green, Over, & Pyne, 1997; Klauer, 1999; Laming, 1996; Oberauer, Wilhelm, & Diaz, 1999), and some of these are still under debate. However, it is notable that the theory allows us to account for various experimental results on the selection task in a single theoretical framework. Theoretical revision and expansion, as well as experimental studies, have followed the initial proposals (e.g., Chater & Oaksford, 1999; Oaksford & Chater, 1996, 1998; Oaksford, Chater, & Grainger, 1999; Oaksford, Chater, & Larkin, 2000).

The ODS model still has some problematic features. First, a questionable assumption is made in formalizing prior hypotheses in terms of probabilities. Correcting this problem (see later) changes the model's predictions, especially on $\neg p$ card selections.

Second, the ODS model does not make quantitative predictions. In general, any model of the selection task should explain the differences in card selection frequencies. The ODS model predicts these differences based on the expected information gain of turning each card, which is calculated from the subjective probabilities of the antecedent (p) and the consequent (q) of the rule ("if p then q"). Though the expected information gains and the proportions of cards selected are continuous values, the ODS model assumes the relationship to be merely monotonic, and it predicts only the order of card selection frequencies. Analysis of the order of card selection frequencies is necessary, but not sufficient.

Third, there are some data that do not seem to be explained by the ODS model. In Kirby's (1994) experiment, it was observed that $\neg p$ selections increased as the subjective probability of the antecedent, $P(p)$, increased, whereas the ODS model only predicts that the selection frequency of the $\neg p$ card should always be the lowest. A similar tendency was also observed in Experiment 2 of Oaksford et al. (1999). In that study, the selection frequency of the $\neg p$ card became comparatively high (.42) when the probability of p was high (.83). Likewise, the results of some experiments (e.g., Oaksford et al., 1999; Oberauer et al., 1999), which examined the effects of $P(p)$ and $P(q)$, did not uniformly verify the ODS model's predictions. Moreover, Evans and Over (1996b) pointed out that the ODS model cannot explain the result of Pollard and Evans (1983), which was intended to manipulate participants' prior beliefs. In contrast to the prediction of the ODS model that prior belief has very little effect, the results showed that $\neg q$ selection increased when participants had low belief in the rule.

Finally, the ODS model does not deal with individual differences. It seems unlikely that all participants solve the selection task in the same way (see, Stanovich, 1999; Stanovich & West, 1998).

In this paper, the ODS model is revised as follows. (1) The questionable assumption about the formalization of hypotheses is corrected. This improves the fit of the model to the data, especially for $\neg p$ selections. (2) A method of parameter estimation using a "selection tendency function" is incorporated into the model. This enables the model to account quantitatively for the results on an arbitrary task by any group of participants. (3) A detailed method to distinguish and to classify individual differences or strategies is proposed. This is applied in two experiments investigating the adequacy of the revised model. The new quantitative model is called the Quantitative Optical Data Selection or QODS model.

BAYESIAN ANALYSIS OF THE SELECTION TASK RE-EXAMINED

Formalizing Wason's selection task in terms of probability enables participants' behaviour on the task to be modelled using Bayesian hypothesis testing. According to the Bayesian approach, rules (in the selection task) are considered as hypotheses to be tested, and a hypothesis can be tested by comparing it to alternative hypotheses. As with the ODS model, it is assumed that two hypotheses are involved in the selection task. One hypothesis, r, is the rule "if p then q" (e.g., a person believes that "if it is a raven [p], then it is black [q]"), corresponding to material implication $p \rightarrow q$ in standard logic. In terms of probability, it is represented by the joint probability distribution shown in Table 1 (left). In this table, x_0 and y_0 represent $P(p \mid r)$ and $P(q \mid r)$, or the probability of the antecedent (i.e., being a raven) given this hypothesis (i.e., in the case that the belief is true) and the probability of the consequent (i.e., being black) given this hypothesis, respectively. Here, it is presupposed that the probability of the consequent is greater than the probability of the antecedent:

$$x_0 \leq y_0 \tag{1}$$

As the ODS model assumed, the alternative hypothesis to which r is compared is assumed to be "p and q are independent"—that is, $\bar{r}$. This hypothesis is expressed by the joint probability distribution shown in Table 1 (right). In this table, x_1 and y_1 denote $P(p \mid \bar{r})$ and $P(q \mid \bar{r})$, respectively. As Evans and Over (1996b) pointed out, the alternative hypothesis to r may be a problematic issue. However, as discussed later, the independence hypothesis, $\bar{r}$, is the simplest and most sufficient hypothesis for explaining the data.

In the Bayesian framework, data update prior beliefs, which are represented by the joint probability distributions r and $\bar{r}$. For Bayesian updating, these probability distributions should be unified, with each weighted by the degree of prior belief, represented by a subjective probability. Here, the relationships between corresponding probabilities of p and q in r and $\bar{r}$ should be specified—that is, between x_0 and x_1, and between y_0 and y_1. The simplest

TABLE 1
Joint probability distribution for the rule "If *p* then *q*" (*r*) and "*p* and *q* are independent" ($\bar{r}$)

	r		$\bar{r}$	
	q	$\neg q$	q	$\neg q$
p	x_0	0	$x_1 y_1$	$x_1 \bar{y}_1$
$\neg p$	$y_0 - x_0$	$1 - y_0$	$\bar{x}_1 y_1$	$\bar{x}_1 \bar{y}_1$

Note: $x_0 = P(p \mid r)$, $y_0 = P(q \mid r)$; $x_1 = P(p \mid \bar{r})$, $y_1 = P(q \mid \bar{r})$, $\bar{x}_1 = 1 - x_1$, $\bar{y}_1 = 1 - y_1$.

assumptions are made, that all probabilities of p in all hypotheses (r and $\bar{r}$ in this case) are the same, and all probabilities of q are the same:

$$\begin{aligned} x_0 &= x_1 = x \\ y_0 &= y_1 = y \end{aligned} \tag{2}$$

Here x and y represent $P(p)$ and $P(q)$, respectively. Because x and y are considered independent variables in our model, these assumptions are natural and appropriate.

The ODS model, however, assumes $y_0 \neq y_1$. In this model, the probability of q in the absence of p is assumed to be the same in both hypotheses—that is to say, $P(q \mid r, \bar{p}) = P(q \mid \bar{r}, \bar{p})$, or $y_0 - x_0 = (1 - x_1)y_1$, and this assumption implies $y_0 \neq y_1$. It is important to note that this assumption entails a counterintuitive consequence. From Bayes theorem,

$$P(r|q) = \frac{P(q|r)}{P(q)} P(r) = \frac{y_0}{y} P(r)$$

According to the ODS model, $P(r|q) \neq P(r)$ is derived because $y_0 \neq y$. As a result, with regard to entropies (i.e., uncertainties in the truth) of the rule ($R = \{r, \bar{r}\}$), it can be shown that

$$H(R|q) \neq H(R)$$

This means that the entropy of R is affected by whether or not the q card is presented. In other words, one would obtain some information about the truth or falsity of the rule only by looking at a q card face up on the desk. If it is felt that this consequence is strange, then it must be concluded that the ODS model's assumption, $y_0 \neq y_1$, is not appropriate.[1]

MODELLING THE TENDENCY TO SELECT A CARD

Generally, turning over a card provides some amount of information about the truth or falsity of the hypothesis. If one turns over the p card, and the flip side is $\neg q$ then this datum refutes the hypothesis "$r: p \rightarrow q$" and reduces uncertainty completely. On the other hand, if the flip side is q, then this datum supports hypothesis r and will increase the degree of belief in hypothesis r. This means that uncertainty is reduced by some amount.

It is assumed that one is more likely to turn over a card when the expected amount of uncertainty reduction is increased. It is also assumed that this relationship is not only monotonically increasing as the ODS model supposed, but is expressed by a specific function as discussed later on.

Scaled information measure

In terms of probability, the degree of uncertainty reduction is defined as the difference in entropies before and after obtaining the information. For example, the expected reduction of uncertainty about the validity of the rule that is obtained by turning over the vowel card is defined as the difference of the entropies—the entropy when the vowel card is facing upwards and the entropy when an even or an odd number is exposed on the other side of the vowel card.

[1] McKenzie and Mikkelsen (2000) argued that whether assumptions about the marginal probabilities—that is, $P(p)$ and $P(q)$—are reasonable might depend on the context (p. 366).

For formalization, let p, $\overline{p}$, q, and $\overline{q}$ now denote a vowel, a consonant, an even number, and an odd number, respectively. In addition, let $P = \{p, \overline{p}\}$ and $Q = \{q, \overline{q}\}$. Therefore, the expectation of uncertainty reduction by turning the vowel card can be written as $H(R\,|\,p) - H(R\,|\,Q, p)$. This is equal to the *expected information gain* and is known to be equal to the *mutual information* (see, e.g., Cover & Thomas, 1991) of the rules and the events on the numerical side conditioned on the vowel card, $I(Q; R\,|\,p)$ (see Appendix). It is abbreviated to $I(p)$, and the expected information gain by turning over a card $x \in P \cup Q$ is defined as

$$I(x) \triangleq H(R|x) - H(R|Y, x) \tag{3}$$

where $Y = P, Q$ and $Y \not\ni x$ (e.g., when x is p or $\overline{p}$, Y is Q). The equations to calculate each $I(x)$ using $P(p)$, $P(q)$, and $P(r)$ are shown in the Appendix.

As the ODS model suggests, the choice of a card in the selection task should be regarded as competitive. Suppose one card (e.g., the p card) is extremely prominent in comparison with other cards, then from the theoretical information point of view it will have a greater chance of being selected. On the other hand, if all four cards are indistinguishable from each other, then every card will have a certain chance of being chosen. Incorporating this view from the ODS model, the expected information gain of card x, $I(x)$, is scaled by the sum of the expected information gains of all cards. This is abbreviated to *scaled information measure*. It is indicated by $Is(x)$ and defined as

$$Is(x) \triangleq \frac{I(x)}{\sum_{x_i \in P \cup Q} I(x_i)} \tag{4}$$

Here, the index i runs over the four cards. As a technical issue, dividing $I(x)$ for each card by the sum (not the mean as in ODS) of these values makes each scaled value vary between 0 to 1 and makes it easier to interpret.

When scaling information, the ODS model requires an additional operation in order to account for the participant's selection of the $\neg p$ card: adding a "small" fixed constant (.10) to all the $I(x)$s. We cannot find any satisfactory reason why such a fixed constant should be added. Moreover, this value is not at all small. For example, when $P(p) = P(q) = .10$, and $P(r) = .50$, where the *rarity assumption* (Oaksford & Chater, 1994) is made, the expected information gain of the p, $\neg p$, q, and $\neg q$ cards, respectively, is .85, .00, .21, and .05.[2] In this standard case, comparing the value for the $\neg q$ card (.05) or the q card (.21), .10 is not a small value. In sum, adding any fixed constant is not justifiable.

The ODS model predicts that the selection frequency of the $\neg p$ card is always the lowest. On the other hand, the QODS model predicts that the $\neg p$ card will be selected with comparatively high frequency where $P(p)$ and $P(q)$ are both high. For example, when $P(p) = .75$ and

[2]The scaled expected information gain (EIG) values that appeared in Table 17.2–17.5 in Oaksford and Chater (1998, pp. 381–389) and in Table 1 in Klauer (1999, p. 217) are miscalculated. They assumed that the posterior belief in the hypothesis should be $P(r_k)P(q_j|r_k, p_i)/P(q_j|p_i)$ for p_i (i.e., p and $\neg p$ cards), and $P(r_k)P(p_i|r_k, q_j)/P(p_i|q_j)$ for q_j (i.e., q and $\neg q$ cards), respectively, where $\{p_i\} = \{p, \overline{p}\}$, $\{q_j\} = \{q, \overline{q}\}$, and $\{r_k\} = \{r, \overline{r}\}$ (M. Oaksford, personal communication, October 1999; in thankful acknowledgement to Mike Oaksford for providing a computer program on which their data were based). However, the posterior probability of the hypothesis given both p_i and q_j (q_j on the reverse side of p_i, or p_i on the reverse side of q_j), should be defined by $P(r_k|p_i, q_j)$. In fact, under the assumption described by Equation 2 in the text, these three equations on the posterior agree with each other.

$P(q) = .77$ (see later for the source of these values), $Is(\overline{p})$ (=.48) is greater than $Is(p)$ (=.11) and $Is(q)$ (=.07).

Although the ODS model yields $Is(x)$ for each card as a continuous value, Oaksford and Chater (1994) analysed only the ordering of the card selection frequencies. Oaksford, Chater, Grainger, and Larkin (1997) argued, however, that if the difference between the scaled information measures (they called it the scaled expected information gain, denoted by $SE[I_g(x)]$) is small, then discriminability is lessened, and card selection becomes ambiguous as a result (p. 444). The logical implication of this statement is that if there exists a clear difference in the frequencies of card selection among the four cards, the scaled information measured $Is(x)$ of these cards would differ correspondingly.

For instance, where $P(p) = .20$ and $P(q) = .21$ (where rarity is maintained), the scaled information measures, $Is(x)$, of the p, $\neg p$, q, and $\neg q$ cards, is computed as .60, .07, .18, and .15, respectively, by the ODS model,[3] but as .47, .07, .38, and .08, respectively, by the QODS model. With regard to the ODS model, it can be seen that not only is the $Is(q)$ value still low, but also that the difference between $Is(q)$ and $Is(\overline{q})$ is quite small, as argued by Oaksford et al. (1997, pp. 443–444) in the context of the reduced array selection task (RAST; Johnson-Laird & Wason, 1970). Although it is true that the order of $Is(x)$ values calculated by the ODS model does coincide with the experimental data, this model cannot predict card selection frequency. The difference in selection rates between q card and $\neg q$ card is usually high; the former is known to show quite high selection rates in standard abstract tasks, and the latter is one of the logically correct solutions that is rarely selected. Neglecting this fact and analysing only selection orders is not sufficient. The QODS model manipulates quantities of information and enables the prediction of card selection frequencies. The method and some assumptions are detailed in the next section.

Modelling alternative hypotheses

There are assumed to be only two *artificial* hypotheses in the ODS model: the dependence hypothesis r, "if p then q", where $P(q \mid r, p) = 1$; and the independence hypothesis $\overline{r}$, "p and q are independent", where $P(q \mid \overline{r}, p) = P(q \mid \overline{r})$. One criticism of the theory was that these assumptions are not general enough (Evans & Over, 1996a), and some possibilities for extension have been shown, such as the incompleteness of the relationship of dependency (Chater & Oaksford, 1999; Green & Over, 1997; Oaksford & Chater, 1998; Oaksford et al., 2000; Over & Jessop, 1998), or a plurality of alternative hypotheses (Green & Over, 1997; Green et al., 1997). Here, we examine how these extensions strengthen the model and how they change its behaviour.

As Oaksford and Chater (1998) argued, we rarely use rules that admit no exceptions in our everyday life. Using $u = P(q \mid r, p)$, a constraint on the probabilistic relationship between p and q, which allows exceptions, can be expressed as $y_0 \leq u \leq 1$, as opposed to $u = 1$, which rejects exceptions. Therefore, an *index of conditional dependency* that indicates the strength of the dependence of q upon p can be defined by

[3]For comparison, the scaled information values in the ODS model were calculated by dividing by the sum, not the mean, of the $I(x)$s.

$$c = \frac{u - y_0}{1 - y_0}$$

where $c = 1$ means complete dependency, and $c = 0$ means complete independency. The joint probability distribution of the complete dependency rule shown in Table 1 (left) can be extended and redefined as r_0 shown in Table 2 (left), which is essentially the same analyses as Oaksford et al. (2000). The QODS model can be constructed with the hypothesis of Table 2 (left) instead of Table 1 (left).

As an alternative hypothesis to the rule "if p then q", one can consider the opposite rule "if p then $\neg q$". This state of affairs can be realized by a model involving three hypotheses: r_0, r_1, and r_2. Here r_0 and r_1 correspond to r and $\bar{r}$, respectively, and r_2 corresponds to the material implication, $p \rightarrow \neg q$. The joint probability distribution of the rule r_2 is shown in Table 2 (right). Here is it presupposed that the probability of the antecedent, $x_2 = P(p \mid r_2)$, plus the probability of the consequent, $y_2 = P(q \mid r_2)$, is less than or equal to 1:

$$x_2 + y_2 \leq 1$$

One can then consider another version of the QODS model that involves hypotheses $R = \{r_0, r_1, r_2\}$.

Not only can many kinds of model be considered to represent the selection task, but the manner in which prior probabilities are assigned is controversial in Bayesian statistics (e.g., Berger, 1985; Earman, 1992). Fortunately, however, as shown in Table 3, priors in the QODS model generally have very weak effects on changing each $Is(x)$. In this table, scaled information measures for three kinds of combination of subjective probabilities, $P(p)$, $P(q)$, and various types of prior, are shown. The kinds of combination of $P(p)$ and $P(q)$ are L–L, L–H, and H–H (L: LOW, H: HIGH); H–L is rejected because this violates the model's presupposition $P(p) \leq P(q)$ derived from Equations 1 and 2. Here, it is assumed that LOW is .23 for p and .25 for q because these are the values that provide the maximum likelihood when estimating the selection tendency function (STF) detailed in the next section; HIGH is .75 for p and .77 for q because .75 = 1 – .25 and .77 = 1 – .23. There are three types of hypothesis compounds in Table 3: complete dependency (r) versus independence ($\bar{r}$), incomplete dependency (r_0, $c \neq 1$) versus independence (r_1), and the competition between complete dependency (r_0, $c = 1$), independence (r_1), and complete opposite dependency (r_2). On the whole, except for the combination H–H with the low prior, $P(r_0) = .10$ or .33, in the three-hypothesis model, little effect can be observed. In general, it is clear that these different hypothesis compounds have little effect on $Is(x)$.

TABLE 2
Joint probability distribution for the extended rule "If p then USUALLY q" (r_0) and another alternative rule "If p then NOT q" (r_2)

	r_0		r_2	
	q	$\neg q$	q	$\neg q$
p	$x_0 u$	$x_0 \bar{u}$	0	x_2
$\neg p$	$y_0 - x_0 u$	$\bar{y}_0 - x_0 \bar{u}$	y_2	$1 - x_2 - y_2$

Note: $x_0 = P(p \mid r_0), y_0 = P(q \mid r_0), u = P(q \mid r_0, p), \bar{u} = 1 - u$; $x_2 = P(p \mid r_2), y_2 = P(q \mid r_2)$.

TABLE 3
Values of the scaled information measure for various prior and subjective probabilities

Type	$P(p)$	$P(q)$	$P(r_0)$	c	$P(r_1)$	$P(r_2)$	$Is(p)$	$Is(\overline{p})$	$Is(q)$	$Is(\overline{q})$
L–L	.23	.25	.80	1.00	.20	–	.52	.07	.29	.13
			.50	1.00	.50	–	.49	.07	.32	.11
			.50	0.80	.50	–	.47	.07	.36	.09
			.50	0.50	.50	–	.47	.06	.39	.08
			.20	1.00	.80	–	.49	.07	.34	.10
			.80	–	.10	.10	.51	.06	.31	.12
			.33	–	.33	.33	.49	.06	.37	.08
			.10	–	.10	.80	.50	.04	.41	.05
L–H	.23	.77	.80	1.00	.20	–	.49	.01	.01	.49
			.50	1.00	.50	–	.48	.02	.02	.48
			.50	0.80	.50	–	.48	.02	.02	.48
			.50	0.50	.50	–	.47	.03	.03	.47
			.20	1.00	.80	–	.48	.02	.02	.48
			.80	–	.10	.10	.45	.05	.05	.45
			.33	–	.33	.33	.43	.07	.07	.43
			.10	–	.10	.80	.41	.09	.09	.41
H–H	.75	.77	.80	1.00	.20	–	.13	.29	.07	.52
			.50	1.00	.50	–	.11	.32	.07	.49
			.50	0.80	.50	–	.09	.36	.07	.47
			.50	0.50	.50	–	.08	.39	.06	.47
			.20	1.00	.80	–	.10	.34	.07	.49
			.80	–	.10	.10	.13	.33	.10	.45
			.33	–	.33	.33	.21	.26	.18	.35
			.10	–	.10	.80	.27	.21	.25	.27

Thus, it can be concluded that assuming the complete dependency versus independence model and that $P(r) = .50$ is reasonable and sufficient. In all subsequent analyses, these are assumed except when explicitly noted.

Selection tendency function (STF)

In the QODS model, it is assumed that the more informative a card will be when turned over, the more likely it is that it will be selected. For instance, if the scaled information measure of a card is extremely prominent (e.g., .80) compared with others, then its probability of being selected may be high (e.g., .95). On the other hand, even when a card provides almost no information, it may have a certain probability of being chosen. The relation between the informativeness of a card and the tendency for it to be turned over is defined by a *selection tendency function* (STF), f.

If the subjective probabilities $P(p)$, $P(q)$, and $P(r)$ in a Wason's selection task with the rule "if p and q", are given, the scaled information measure $Is(x)$ can be calculated for each card x from Equations 3 and 4. Each value of $Is(x)$ can then be matched against a card selection tendency (i.e., a probability of being selected), $T(x)$, with the STF. The card selection tendency

$T(x)$ will be realized as $R(x)$, which is the card selection ratio (see later) for the target group of participants. In other words, $T(x)$ is an estimate of $R(x)$.

How to estimate the subjective probabilities $P(p)$ and $P(q)$ from the observed card selection rates can now be considered. If $R(x) = d/N$ is given, let d be the observed number of selections of card x, and let N be the total number of participants. Because $R(x)$ is described by a binomial distribution $B(N, T)$, the likelihood function D, which is the vector of ds, is

$$like(D) = \prod_i \binom{N}{d_i} T(x_i)^{d_i} [1 - T(x_i)]^{N-d_i} \tag{5}$$

To maximize this function, consider its natural logarithm (omitting the constant):

$$L(D) = \sum_i d_i \log T(x_i) + \sum_i (N - d_i) \log[1 - T(x_i)] \tag{6}$$

The pair of values of $P(p)$ and $P(q)$ that maximize Equation 5 will maximize Equation 6. These values are called estimated values of $P(p)$ and $P(q)$, or *maximum likelihood estimates* (MLEs) for the data $R(x)$.

To obtain estimates, the STF, f, which maps $Is(x)$ onto $T(x)$ with a monotonic relationship, must be specified. The range of $T(x)$ should be $0 \le T(x) \le 1$ because $T(x)$ is a probability, but $T = 0$ and $T = 1$ are not desirable in view of the estimation process. From these requirements, f is assumed to be a logistic function:

$$T(x) = f[Is(x)] = \frac{1}{1 + e^{-Z(x)}} \tag{7}$$

where $Z(x) = a + bIs(x)$, which is called the logit of $T(x)$. In this function, parameter b defines the inclination of the slope, which provides the *contrast* between any two points with different scaled information measures, and parameter a defines the position of the curve, which provides *sensitivity* to the information.

Parameter estimation of the STF

Using experimental data for $R(x)$, the parameters a and b of the STF in Equation 7 can be estimated by the maximum likelihood method, under the assumption that a and b are independent of $P(p)$, $P(q)$, and $P(r)$. The experimental data used in the estimation procedure were chosen from studies reporting individual card selection frequencies, which were cited for comparison in Oaksford and Chater (1994). The data were carefully screened because ideally the values of $P(p)$ and $P(q)$ should be the same for each experiment. Any studies with *violation instructions*, with *deontic rules*, with nonstudent participants, and with nonstandard tasks (e.g., with thinking aloud) were rejected so that the experimental conditions would yield homogeneous data.[4] Selected data are shown in the upper part of Table 4.

Equation 6 is a nonlinear function with four variables, a, b, $P(p)$, and $P(q)$. To maximize it globally, a numerical approach based on the Newton–Raphson method with a grid refinement

[4]Strictly speaking, the STF differs for each experiment because different participant groups bring different knowledge or background beliefs to the task; furthermore, it differs for each individual. However, it is still useful to group various data for the selection task by estimating an STF by carefully averaging possible differences.

TABLE 4

Selection frequency data from studies on abstract selection tasks, maximum likelihood estimates (MLEs) of $P(p)$, $P(q)$, and prediction of the model

				Data				*MLE*				*Model*			
No.	*Study*	*Experiment, condition*	*N*	$R(p)$	$R(\bar{p})$	$R(q)$	$R(\bar{q})$	$\hat{P}(p)$	$\hat{P}(q)$	G^2	*P*	$\hat{R}(p)$	$\hat{R}(\bar{p})$	$\hat{R}(q)$	$\hat{R}(\bar{q})$
1	Wason (1968)	expt 1, expt group	18	1.00	.17	.78	.28	.18	.21	5.9	.05	.91	.13	.68	.16
2	Wason (1968)	expt 1, control group	16	1.00	.06	.69	.13	.04	.05	1.3	.53	.97	.09	.66	.09
3	Wason (1968)	expt 2	26	1.00	.12	.50	.12	.06	.10	1.2	.56	.98	.09	.48	.10
4	Evans and Lynch (1973)	a/a	24	.88	.08	.50	.33	.28	.34	1.5	.47	.89	.16	.48	.30
5	Manktelow and Evans (1979)	expt 1, abstract, a/a	24	.96	.13	.63	.33	.24	.29	2.7	.26	.91	.14	.55	.23
6	Manktelow and Evans (1979)	expt 2, abstract, a/a	24	.83	.17	.67	.25	.27	.29	0.1	.96	.83	.18	.66	.23
7	Manktelow and Evans (1979)	expt 3, abstract, a/a	16	.88	.38	.56	.38	.35	.38	2.8	.25	.78	.23	.53	.35
8	Manktelow and Evans (1979)	expt 4, abstract, a/a	16	.69	.00	.63	.13	.25	.25	9.1	.01	.83	.17	.74	.19
9	Manktelow and Evans (1979)	expt 5, abstract, a/a	16	.88	.13	.81	.06	.12	.13	0.6	.73	.89	.11	.83	.12
10	Griggs and Cox (1982)	expt 1, abstract, a/a	32	1.00	.06	.63	.06	.04	.06	2.1	.35	.98	.08	.60	.09
11	Griggs (1984)	expt 1, t/f, non-mem	25	.84	.16	.52	.08	.23	.26	4.9	.09	.90	.14	.61	.21
12	Chrostowski and Griggs (1985)	expt 2, t/f, non-mem	60	.97	.07	.78	.08	.04	.05	0.8	.68	.95	.09	.77	.09
13	Hoch and Tschirgi (1985)	bachelor, standard	25	.88	.24	.60	.40	.32	.36	2.0	.37	.82	.20	.53	.32
14	Valentine (1985)	t/f, a/a	24	.83	.13	.58	.25	.28	.31	0.5	.76	.86	.17	.60	.25
15	Yachanin (1986)	expt 2, t/f, abstract	20	.95	.20	.70	.30	.24	.27	2.6	.27	.88	.15	.62	.22
16	Beattie and Baron (1988)	expt 1, 4-card, a/a	18	.94	.06	.56	.06	.04	.05	1.4	.51	.98	.08	.59	.09
17	Oaksford and Stenning (1992)	expt 2, abstract, a/a	24	.79	.25	.63	.21	.31	.32	0.6	.73	.78	.21	.65	.26
18	Oaksford and Stenning (1992)	expt 3, coloured shape, a/a	24	.63	.29	.63	.29	.36	.36	0.3	.85	.66	.29	.66	.29
19	Oaksford and Stenning (1992)	expt 3, vowel–even, a/a	24	.96	.17	.71	.29	.23	.25	3.1	.21	.90	.14	.63	.20
20	Oaksford and Stenning (1992)	expt 3, control, a/a	24	.88	.17	.50	.17	.25	.29	1.3	.53	.90	.15	.55	.24
21	Evans and Lynch (1973)	a/n	24	.92	.04	.08	.58	.22	.61	2.6	.27	.96	.10	.12	.63
22	Manktelow and Evans (1979)	expt 1, a/n	24	1.00	.08	.21	.75	.23	.64	4.6	.10	.95	.10	.12	.69
23	Manktelow and Evans (1979)	expt 2, abst, a/n	24	.96	.04	.33	.75	.31	.55	7.6	.02	.91	.13	.17	.66
24	Valentine (1985)	t/f, a/n	24	.92	.21	.54	.38	.30	.35	1.3	.51	.87	.17	.49	.32
25	Beattie and Baron (1988)	expt 1, 4-card, a/n	18	1.00	.06	.06	.61	.02	.95	1.0	.60	.98	.08	.08	.59
26	Oaksford and Stenning (1992)	expt 2, abstract, a/n	24	.83	.25	.17	.67	.39	.54	1.1	.56	.82	.18	.22	.69
27	Oaksford and Stenning (1992)	expt 3, coloured shape, a/n	24	.58	.21	.63	.29	.36	.36	1.6	.45	.66	.29	.66	.29
28	Oaksford and Stenning (1992)	expt 3, vowel–even, a/n	24	.88	.13	.58	.33	.28	.32	0.9	.64	.87	.17	.54	.28
29	Oaksford and Stenning (1992)	expt 3, control, a/n	24	.92	.08	.33	.46	.29	.42	1.1	.58	.92	.15	.30	.43

30	Evans and Lynch (1973)	n/n	24	.54	.46	.29	.75	.54	.59	0.7	.71	.49	.41	.27	.72
31	Manktelow and Evans (1979)	expt 1, abst, n/n	24	.83	.21	.54	.67	.39	.46	6.1	.05	.78	.23	.36	.51
32	Manktelow and Evans (1979)	expt 2, abst, n/n	24	.71	.33	.38	.58	.44	.49	0.4	.84	.68	.29	.36	.56
33	Valentine (1985)	t/f, n/n	24	.88	.25	.54	.38	.33	.37	1.0	.60	.83	.20	.50	.34
34	Oaksford and Stenning (1992)	expt 2, abst, n/n	24	.88	.25	.54	.42	.34	.38	1.4	.49	.82	.20	.48	.36
35	Oaksford and Stenning (1992)	expt 3, coloured shape, n/n	24	.67	.29	.63	.21	.36	.36	0.9	.63	.66	.29	.66	.29
36	Oaksford and Stenning (1992)	expt 3, vowel-even, n/n	24	.83	.13	.63	.38	.31	.34	1.9	.38	.83	.19	.57	.29
37	Oaksford and Stenning (1992)	expt 3, control, n/n	24	.71	.17	.38	.54	.41	.47	1.2	.55	.75	.25	.36	.53
38	Evans and Lynch (1973)	n/a	24	.58	.29	.58	.42	.40	.41	0.8	.67	.64	.32	.58	.35
39	Manktelow and Evans (1979)	expt 1, abst, n/a	24	.79	.29	.71	.50	.36	.38	5.3	.07	.73	.26	.58	.33
40	Manktelow and Evans (1979)	expt 2, abst, n/a	24	.83	.29	.79	.29	.28	.29	2.9	.23	.77	.20	.73	.22
41	Valentine (1985)	t/f, n/a	24	.75	.29	.67	.17	.32	.32	1.5	.47	.72	.24	.70	.25
42	Oaksford and Stenning (1992)	expt 2, abst, n/a	24	.58	.33	.71	.42	.40	.40	1.8	.41	.61	.33	.61	.33
43	Oaksford and Stenning (1992)	expt 3, coloured shape, n/a	24	.58	.33	.63	.17	.36	.36	3.0	.23	.66	.29	.66	.29
44	Oaksford and Stenning (1992)	expt 3, vowel-even, n/a	24	.88	.08	.75	.25	.21	.22	1.8	.40	.88	.14	.73	.17
45	Oaksford and Stenning (1992)	expt 3, control, n/a	24	.71	.29	.54	.38	.39	.40	0.0	.99	.70	.28	.54	.37
46	Kirby (1994)	expt 1, S	45	.78	.29	.60	.49	.38	.41	4.1	.13	.75	.25	.49	.40
47	Kirby (1994)	expt 1, L	44	.55	.48	.41	.73	.53	.56	3.7	.13	.50	.42	.31	.66

Note: The data in the upper part of this table (1–20) are used for the STF estimation detailed in the text. G^2 indicates the likelihood ratio of the estimated values, which is a measure of goodness-of-fit, and P indicates the exact probability, or critical value of α to reject the model.

In the "Experiment, condition" column, a/a means affirmative antecedent and affirmative consequent condition, a/n means affirmative antecedent and negative consequent condition, n/a means negative antecedent and affirmative consequent condition, n/n means negative antecedent and negative consequent condition.

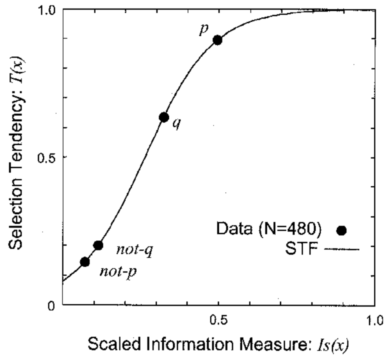

Figure 1. The selection tendency function (STF).

algorithm was executed by a computer program (Mathematica 4.0, 1999; Global Optimization 3.0, 1998). The STF was determined with the parameters $a = -2.43$ and $b = 9.27$ as shown in Figure 1. In this case, $P(p) = .23$ and $P(q) = .25$.

META-ANALYSIS USING THE MODEL

The estimated parameters of the STF, a and b, are regarded as semi-fixed in general, whereas the estimated subjective probabilities, $P(p)$ and $P(q)$, can differ in every experiment as the probabilistic information varies. The purpose of this section is to evaluate how well the model fits the data on the selection task by seeking the best fitting values of $P(p)$ and $P(q)$. These values can then be assessed for whether they conform to those predicted.

Method

Given the subjective probabilities $P(p)$ and $P(q)$, the expected information gain $I(x)$ and the scaled information measure $Is(x)$ of each of the four cards can be calculated. From this, the selection tendency $T(x)$ that corresponds to the selection frequency of each card is derived using the STF $f(x)$ (see Figure 2). Conversely, participants' potential subjective probabilities can be estimated from the experimental data

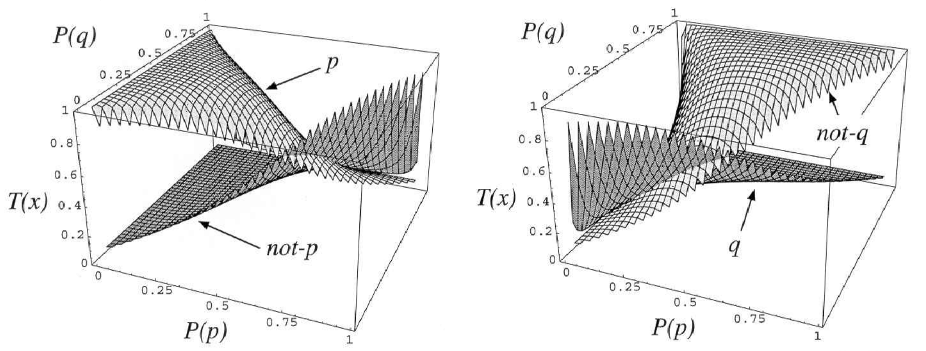

Figure 2. Three-dimensional surface plots of the selection tendencies of card p and $\neg p$ (left) and cards q and $\neg q$ (right) as functions of $P(p)$ and $P(q)$ when $P(r) = .50$.

on selection frequencies.[5] To estimate $P(p)$ and $P(q)$, the same method as that for estimating the STF described earlier can be used. The only difference is on the number of parameters, because a and b, have already been determined. The purpose is to find the values of $P(p)$ and $P(q)$ that provide the global maximum of Equation 6 given $R(x)$ for each of the four cards, which is regarded as $T(x)$. Using this method, the latent subjective probabilities $P(p)$ and $P(q)$ of some experiments in the literature were estimated.

The target data were restricted to *nondeontic* abstract tasks. The reason for selecting nondeontic tasks is that in a deontic task, people can also use some other kind of measure than the amount of information (e.g., the subjective utility proposed by Kirby, 1994, and Oaksford & Chater, 1994). The reason for the restriction to abstract tasks is to compare the subsequent experiments where probabilistic information is experimentally varied. The target tasks were abstract standard selection tasks (which were used to identify the STF), negations paradigm abstract tasks (see later), and tasks with probabilistic information as used in Kirby (1994), which are called *probabilistic tasks* in this paper.

Results and discussion

All of the results of the estimations are shown in Table 4, and a scatterplot is presented in Figure 3. Table 4 reveals good fits to the data. The model could only be rejected for one experiment, assuming a $P < .01$ criterion (.05 is regarded as unreasonably large for model rejection, Read & Cressie, 1988). The results are described according to the kind of task.

Standard abstract task

The results of estimating $P(p)$, $P(q)$ for the experimental data using standard abstract selection tasks are plotted by circles (single or double) in Figure 3. All of these points are distributed within the area (approximately) $P(p) < .4$ and $P(q) < .4$. This result supports the rarity assumption of Oaksford and Chater (1994).

The estimated values all lie in the vicinity of the line $P(q) = P(p)$. This result corresponds to the participants' tendency to a biconditional interpretation. If the rule given is biconditional (i.e., the rule is "if and only if p then q"), $y_0 - x_0$ in Table 1 (left) is always equal to zero; therefore, $x = y$, or $P(p) = P(q)$. In the neighbourhood of the line $P(q) = P(p)$, the conditional rule is interpreted more biconditionally. The results mean that participants in all of these studies tend to interpret the conditional rule biconditionally. It has been argued that people have a tendency to interpret a conditional sentence as a biconditional one (e.g., Geis & Zwicky, 1971; Taplin, 1971), except for some cases of relatively familiar categorical relationships (e.g., "If it is a raven, then it is black"). Generally speaking, when there is no information of $P(p)$ and $P(q)$ it is assumed that the properties that figure on the conditionals are bidirectional. This is labelled the *biconditionality assumption*.

As Platt and Griggs (1995) pointed out, rules in the abstract selection task can be classified into two types: Categorical rules (e.g., "If a card has a vowel on one side, then it has an even number on the other side"), and specific-instance rules (e.g., "If there is an A on one side of the

[5]The estimated value of either $P(p)$ or $P(q)$ using the method introduced here is not for one person, but for one participant group. In fact, participants retain their own subjective probabilities, but the response to each card, based on these probabilities, is dichotomous: to select or not. We cannot estimate the latent probabilities for one person from a one-shot action, nor can we repeat the same experimental task with the same person. Therefore, it is difficult to identify one person's subjective probabilities. A value estimated for a group of participants can be regarded as a kind of average over all the individuals in the group.

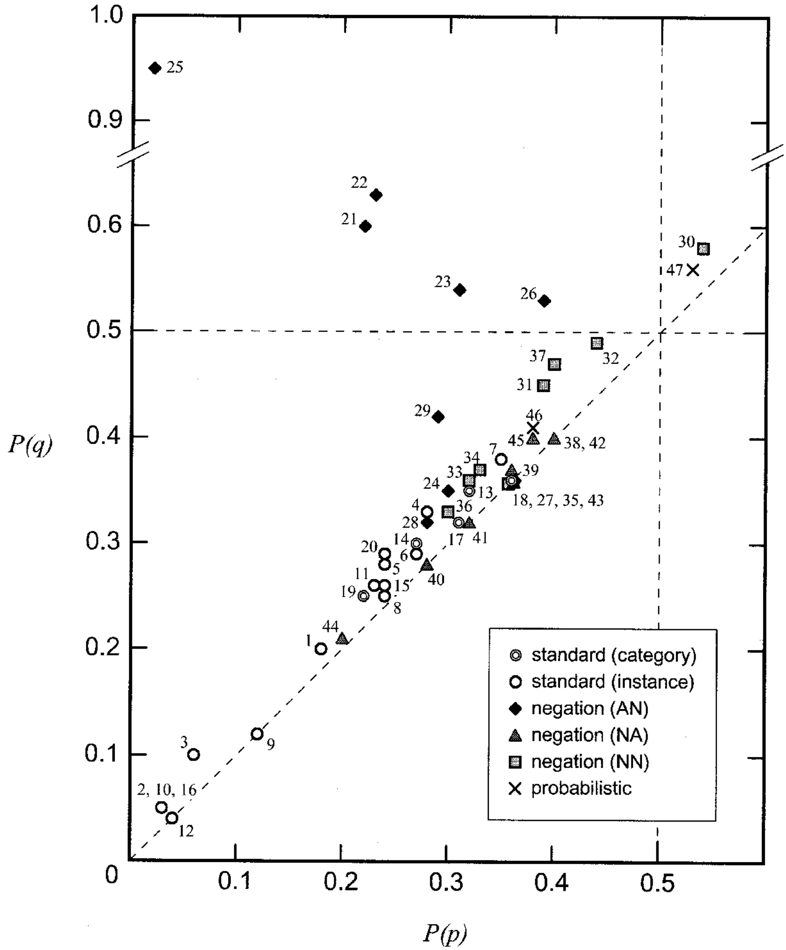

Figure 3. Estimated values of subjective probability $P(p)$ and $P(q)$ for various kinds of abstract selection task. Each figure beside the points corresponds to the sequential number in Table 4.

card then there is a 4 on the other side"). The two kinds of rule were distinguished by the results of the estimation procedure. There were 5 experiments using a categorical rule and 15 experiments using an instance rule. In Figure 3, experiments with a categorical rule are plotted by double circles, and those with an instance rule are plotted by single circles. The results for instance rules are spread over a comparatively wide range, $.03 \leq P(p) \leq .35$, but the results for categorical rules are gathered within a comparatively small area, $.22 \leq P(p) \leq .36$. This difference may be explained as follows. In both cases, no probabilistic information was given in the tasks. Especially in the task with an instance rule, however, the values of subjective probabilities will be highly dependent on each participant's comprehension of the *focal set* (Cheng & Novick, 1991, 1992) or the *contrast class* (Oaksford & Stenning, 1992) for the target materials, which would vary according to participants.

Negations paradigm abstract task

Evans and Lynch (1973) showed that the presence or absence of negated constituents in a rule affects a participant's performance in the selection task. There are four possible patterns for rule in the negations paradigm in the selection task: affirmative antecedent and affirmative consequent (AA); affirmative antecedent and negative consequent (AN); negative antecedent and affirmative consequent (NA); and negative antecedent and negative consequent (NN). The corresponding four cards are referred to as true antecedent (TA), false antecedent (FA), true consequent (TC), and false consequent (FC). Although it might cause some complications as a rule contains negatives, it is assumed here that any negated constitutions contained in the rule are included in the corresponding propositions. In other words, the TA, FA, TC, and FC cards are referred to as the p, $\neg p$, q, and $\neg q$ cards, respectively, regardless of any negated constituents in the rule.

Generally speaking, when the consequent contains a negative constituent (i.e., AN and NN), the number of participants who select the $\neg q$ (logically correct) card increases. Oaksford and Chater (1994) explained these results based on the ODS model and the rarity assumption. If it is assumed that the default values of $P(p)$ and $P(q)$ are low, the probabilities of negation of p and q that are described as $1 - P(p)$ and $1 - P(q)$ must be high. When $P(q)$ is high (the negative is included in q itself in this case), the ODS model predicts a high selection rate of the $\neg q$ card.

Here the probabilities of the antecedent and the consequent in the negations paradigm selection tasks are identified. First, the plotted datum points for the AN rule indicate a wide distribution area in Figure 3. Many of them have low $P(p)$ and high $P(q)$ values. Second, data for the NN rule indicate comparatively high values of $P(p)$ and $P(q)$ (both are generally in the range between .3 and .6). Third, the data for the NA rule show comparatively low $P(p)$ and $P(q)$, and the difference in these values is low. As the results for the NA rule and the standard abstract tasks have common characteristics, they can hardly be distinguished from one another.

This result clearly indicates that negated constituents raise the subjective probability of the events that contain them, and the result is consistent with the rarity assumption. For the NA rule, because a combination of high $P(p)$ and low $P(q)$ is incompatible with the constraint of Equation 1, $P(p)$ was suppressed to about the default level. In consequence, the results were similar to those of the standard AA type abstract task (Oaksford & Chater, 1994).

Examining the subjective probabilities that contain a negated constituent (ignoring the distinction between antecedent and consequent), and ignoring NA, they vary within the range from .30 to .95 ($M = .45$, $SD = .14$, $N = 25$). On the other hand, the range of affirmative events was .03–.38 ($M = .22$, $SD = .11$, $N = 49$). Now, let x be the probability of an affirmative event, and let x' be the probability of the negation of the event. From this result, the relationship of these probabilities would be expressed $x + x' < 1$. This means that the subjective probabilities of negative events are usually lower than those calculated by ordinary probability theory. This property is called *superadditiveness*. The reason and meaning of the property are discussed later in the section General Discussion.

Probabilistic abstract task

Kirby (1994) proposed a theory of the selection task based on subjective utility theory. His model predicted that as $P(p)$ increases, the selection frequency of the $\neg q$ card will increase, and his experiments provided evidence that this is the case. In his experiments, information on $P(p)$ was given by the ratio of the numbers of cards. The base rate of mistakes violating the rule was also given. Although the base rate is closely related to $P(r)$, it may not be regarded as $P(r)$ itself, because of the ambiguity in interpretation of the instructions given in his experiment as pointed out by Over and Evans (1994). Therefore, $P(p)$ and $P(q)$ were estimated with several values of $P(r)$. As $P(r)$ increased from .50 to .99, in the small $P(p)$ condition, the estimated values of $P(p)$ varied from .38 to .36, and $P(q)$ varied from .41 to .37; whereas in the large $P(p)$ condition, $P(p)$ varied from .53 to .56, and $P(q)$ varied from .56 to .56. All of these ranges are small, and so it is clear that the effect of $P(r)$ is very small.

Estimated values of $P(p)$ in the large $P(p)$ condition in Kirby (1994) are obviously higher than those for the small $P(p)$ condition. This result is consistent with the probabilistic information given in the instructions, $P(p) = 1/1001$ in the small $P(p)$ condition, and $P(p) = 1000/1001$ in the large $P(p)$ condition (note that there is a critical argument on this topic in Evans & Over, 1996a). However, estimates of $P(p)$ and $P(q)$ were not so extreme, although the relation $P(p) \ll P(q)$ was maintained. This result is compatible with the concept of the weighting function in *prospect theory* (Kahneman & Tversky, 1979). In this theory, it is proposed that very low probabilities are overweighed, and for all the probabilities, the decision weight satisfies *superadditiveness*.

Furthermore, although there was no explicit description of $P(q)$ at all in the tasks, the relation $P(p) \simeq P(q)$ in all the estimates can be seen. Thus the default relationship between $P(p)$ and $P(q)$ would appear to be $P(p) \simeq P(q)$. In other words, the conditional rule is interpreted biconditionally according to the biconditionality assumption, at least when no explicit information is given.

Evans and Over (1996b) pointed out that the ODS model cannot explain Kirby's (1994) data showing an increase in selections of $\neg p$ cards when $P(p)$ was high. The QODS model, however, can explain these data. According to the model, as $P(p)$ increases, the scaled information measure for $\neg p$ rises, and this causes an increment of selection frequency of the $\neg p$ card.

Evans and Over (1996a) also pointed out that Oaksford and Chater (1994) calibrated $P(r)$ for Kirby's (1994) data without adequate reasons. They claim it is impossible to identify $P(r)$ based on the instructions because of procedural faults. However, as shown earlier, $P(r)$ is powerless to affect the models behaviour. Therefore, this criticism is not very important from a theoretical point of view.

Summary

The results of the meta-analysis using the QODS model raised three important issues. First, the experimental data from the standard tasks and the negations paradigm tasks supported the rarity assumption. The estimates of $P(p)$ and $P(q)$ for affirmative propositions indicated low values, and negations, which flip the truth value, raised these estimates. Second, the results of the negation-free tasks (i.e., standard and probabilistic tasks), and the NA tasks that violate the probabilistic constraints of QODS, supported the biconditionality assumption. Each pair of estimates for $P(p)$ and $P(q)$ indicated similar values.

However, and third, when the task situation violated either rarity or biconditionality the estimated values of $P(p)$ and $P(q)$ were more *moderate* than expected. In the probabilistic tasks, the estimates were not as extreme (e.g., .99 or higher) as those given in the task instruction, and the estimates for negated propositions were higher but not sufficiently so: The sum of the affirmed and the negated estimates was less than 1.00. This superadditiveness is characteristic of the π-function in prospect theory (Kahneman & Tversky, 1979).

EXPERIMENT 1

The most conspicuous quantitative difference between the ODS model and the QODS model is in the prediction of the selection frequency of the $\neg p$ card. The ODS model predicts that the selection frequency of the $\neg p$ card is always the lowest. On the other hand, the QODS model predicts that when both $P(p)$ and $P(q)$ are high, the $\neg p$ card will be selected more than p and q (see Figure 2). In Kirby's (1994) Experiments 1, 2, and 3, as $P(p)$ increased, the selection frequency of $\neg p$ increased. Experiment 1 was designed to investigate how $P(p)$ influences the selection frequency of the $\neg p$ card.

The second purpose of this experiment was to validate the QODS model directly. There seem to be few studies on the abstract version of the selection task that explore the effect of probability on card selection frequencies by systematic variation of the probabilities of the antecedent and the consequent, $P(p)$ and $P(q)$. It is known that facilitation—that is, helping participants select the logically correct cards (p and $\neg q$)—is not a simple matter, especially in the abstract selection task (Platt & Griggs, 1993, 1995). Both models predict that the logical choice will be increased if $P(p)$ is low and $P(q)$ is high (see Figure 2). In an attempt to determine whether the probabilistic information provides facilitation in the abstract selection task, this probability condition, $P(p) = \text{LOW}$ and $P(q) = \text{HIGH}$, was included in the experiment.

Although some experiments (e.g., Oaksford et al., 1999; Oberauer et al., 1999) have manipulated $P(p)$ and $P(q)$, they did not uniformly support the ODS model. The main cause of this may be individual differences among participants (Stanovich & West, 1998). It is obvious that all participants do not perform the task in the same way. However, in a long history of research on human reasoning using Wason's selection task, few studies have analysed participants' strategies. In this experiment, a new method of analysing participants' individual differences from the probabilistic point of view was introduced. The third purpose of this experiment was therefore to examine the validity of this method.

Method

Participants and design

A total of 136 undergraduate students of Ritsumeikan University participated in this experiment. Of these, 3 reported having seen the selection tasks before. The results given here are for the remaining 133 participants. Participants were randomly assigned to one of these groups: 45 of the participants were allocated to the low $P(p)$ and high $P(q)$ condition (Condition L–H), 47 were allocated to the high $P(p)$ and high $P(q)$ condition (Condition H–H), and 41 were allocated to the no probabilistic information or control condition (Condition C).

The tasks were printed in a booklet, and the experiment was administered to the participants as a group. Participants were instructed to perform the tasks according to page sequence and never in the reverse order.

Materials and procedure

To begin, participants read the following instructions (in the case of Condition L–H). This task is called the *probabilistic selection task*.

> There are two sets of Hanafuda[6] cards: one set has a reverse side that is black; the other set has a red reverse side. Now we are going to create a new set of Hanafuda cards by drawing cards from each set. Mr Taniyama was elected to do this work and he claims the following rule was kept while compiling the new set of Hanafuda cards: "If the back of the card is red, then it is a trash card". Examination of the new Hanafuda card set that Taniyama has made shows that there are only a few red cards, and all the others are black; most of the cards are trash cards, and very few are scoring cards.
>
> Arrayed on the desk are 4 cards drawn from the new Hanafuda card set that Taniyama has made. Which card, or cards, need to be turned over in order to determine whether his rule is true or false? Mark the card(s) which you think that need to be turned over. (Red is indicated by grey.)

Next, the following question was asked in order to test the participant's card preference order. This task is called the *preference order task*.

> When the number of cards you can turn over is restricted to only one, which card would you choose to turn over? Write the number 1 on the card. Next, write the number 2 on the card that you would want to turn over second if the number of cards is restricted to two. In the same way, supposing the number is three or four, write a number on each card. You can use the same number for one as for another only if you cannot put them in order.

In Condition H–H, the rule used was: "If the back of the card is black, then it is a trash card". Probabilistic information on the card colour, $P(p)$, was replaced by the sentence: "For the most part cards are black, and the remaining few are red". In Condition C the rule was given without any probabilistic information on the card colour: "If the back of the card is red, then it is a scoring card".

Subsequently, so that the sensitivity of the participants to the probabilistic information in the task could be determined, they were asked to rate, on a 5-point scale, the extent to which they considered each of the following:

1. Did you give consideration to the number of red cards or trash cards in your card selection?
2. Did you give consideration to the certainty of the rule that Mr Taniyama proposed?

The 5-point answer scale ranged from *did not consider at all* (1), to *considered very much* (5).

Furthermore, the participants were asked to evaluate the subjective probabilities, $P(p)$, $P(q)$, and $P(r)$ as percentages (which is called the *probability evaluation task*). It is important that the probability of the rule "if p then q", or $P(r)$, should not be confused with that of q given p, or $P(q\,|\,p)$. The minimize this possible confusion an imaginary person, named Taniyama, was introduced, who claimed the conditional rule to be true. The participants were required to evaluate their degree of confidence in him as an evaluation of $P(r)$.

Finally, participants performed the original type of abstract selection task involving a vowel and an even number.

[6]Hanafuda is a traditional Japanese card game (Japan Publications Inc., 1970). However, any background knowledge of the game is irrelevant to the solution of the task. Hanafuda cards can be divded into two categories: scoring cards and trash cards. A scoring card is described by the Japanese word *ten-fuda*, and trash card is called *kasu-fuda*.

Results and discussion

The proportion of cards selected in each type of task and participant's strategy group, which are described later, are given in Table 5. With regard to the preference order task, a weighted scoring for each card in order of preference was calculated and analysed. The results were, however, very similar to those of the probabilistic selection task, so a report of the preference order data is omitted here.

Original abstract task. The purpose of this task was to check the homogeneity of participants between conditions. Although some effects of transfer from the probabilistic task might exist, a number of studies have failed to discover any transfer effects on the abstract selection task (e.g., Johnson-Laird, Legrenzi, & Legrenzi, 1972; Wason & Shapiro, 1971).

The differences in selection frequencies of each card under the three conditions were not statistically significant, with $\chi^2(2, N = 132) = 0.81, 0.71, 1.60$, and 0.42, for the p, $\neg p$, q, and $\neg q$ cards, respectively, ps > .05 (see Table 5).

Differences between conditions. In Condition L–H, both models predict that the p and $\neg q$ cards will be selected frequently. The result, however, indicates that $R(p) = .64$ and $R(\overline{q}) = .42$, which were, especially for the $\neg q$ card, contrary to both models' predictions.

In Condition H–H, the ODS model predicted a high selection frequency for the p and $\neg q$ cards, whereas the QODS model predicted a high selection frequency for the $\neg q$ and $\neg p$ cards. The result shows $R(p) = .72$, $R(\overline{q}) = .70$, and $R(\overline{p}) = .42$. The ODS model therefore seemed to be supported by these data.

In Condition C, the card selection pattern of the probabilistic task resembles that of the original abstract task. For each card, there was no significant difference between the card selection frequencies: $\chi^2(2, N = 82) = 0.46, 1.30, 0.54$, and 2.40, for the p, $\neg p$, q, and $\neg q$ cards, respectively (ps > .05). This result means that without explicit probabilistic information, the default (low) value of probability is adopted, regardless of task materials.

Grouping participants by strategies. Some participants may think probabilistically in performing the probabilistic task, but others may not. Participants who adopt a probabilistic strategy must be sensitive to information on the probability or frequency of cards, as evaluated in Question 1. Therefore, the rating for Question 1 is called the *probability-consciousness index* (PI). In order to sort participants by their strategy from a probabilistic point of view, all participants were categorized according to their PI: Participants who rated four or more were labelled the *probability-conscious* (PC) group, and the others were called the non-PC group.

With regard to the non-PC group, the patterns of selection frequencies were similar between Conditions L–H and H–H (see Table 5): Both had an enhanced p card selection frequency. No significant differences were detected between the proportions of cards selected, $\chi^2(1, N = 56) = 0.83, 0.65, 0.10$, and 3.19, for the p, $\neg p$, q, and $\neg q$ cards, respectively (ps > .05). The subjective probabilities $P(p)$ and $P(q)$ for the non-PC group were estimated using the method introduced in the previous section: $P(p) = .43$ and $P(q) = .44$ for Condition L–H, and $P(p) = .42$ and $P(q) = .47$ for Condition H–H. These values were similar between conditions. Logically speaking, if we ignore the probabilistic information, Conditions L–H and H–H are the same. These results indicate that the non-PC participants made little use of probabilistic

TABLE 5

Proportion of cards selected in Experiments 1 and 2, MLE, and the prediction of the model for each subgroup of participants

			Data				*MLE*				*Model*			
Expt	*Condition/subgroup*	N	$R(p)$	$R(\overline{p})$	$R(q)$	$R(\overline{q})$	$\hat{P}(p)$	$\hat{P}(q)$	G^2	P	$\hat{R}(p)$	$\hat{R}(\overline{p})$	$\hat{R}(q)$	$\hat{R}(\overline{q})$
1. Probabilistic task	L–H	45	.64	.44	.62	.42	.44	.44	3.1	.22	.56	.38	.55	.39
	H–H	47	.72	.43	.49	.70	.47	.51	10.3	.01	.62	.33	.36	.58
	C	41	.90	.32	.68	.56	.36	.39	17.5	.00	.78	.23	.50	.37
	L–H/PC	21	.57	.48	.71	.43	.44	.44	3.1	.22	.56	.39	.56	.39
	H–H/PC	15	.53	.67	.47	.80	.59	.60	5.5	.06	.38	.54	.30	.67
	L–H/non-PC	24	.71	.42	.54	.42	.43	.44	1.5	.47	.62	.34	.51	.42
	H–H/non-PC	31	.81	.32	.48	.65	.42	.47	5.0	.08	.73	.26	.37	.53
1. Original task	L–H	45	.87	.36	.69	.33	.32	.34	8.1	.02	.77	.22	.63	.28
	H–H	46	.91	.37	.63	.39	.34	.36	12.4	.00	.79	.22	.55	.32
	C	41	.85	.44	.76	.39	.36	.36	15.2	.00	.70	.27	.64	.30
2. Probabilistic task	Whole	82	.78	.29	.60	.56	.40	.43	13.1	.00	.72	.27	.47	.43
	CC	24	.83	.25	.54	.83	.43	.51	12.7	.00	.74	.24	.29	.63
	Non-CC	58	.76	.31	.62	.45	.39	.40	4.2	.12	.70	.28	.54	.37
	PC	27	.89	.26	.52	.59	.37	.44	5.3	.07	.81	.22	.39	.46
	Non-PC	55	.73	.31	.64	.55	.40	.42	8.7	.01	.69	.29	.51	.40
2. Original task	Whole	82	.93	.28	.76	.24	.25	.26	14.1	.00	.85	.16	.70	.20
	CC	24	.92	.29	.63	.17	.25	.27	3.6	.17	.88	.15	.64	.21
	Non-CC	58	.93	.28	.81	.28	.24	.25	13.9	.00	.84	.17	.72	.19
	PC	27	.96	.33	.67	.22	.26	.28	7.8	.02	.87	.16	.62	.23
	Non-PC	55	.91	.25	.80	.25	.24	.25	8.3	.02	.83	.17	.73	.19

Note: In "Cond/subgroup" column, L–H means low $P(p)$ and high $P(q)$ condition, H–H means high $P(p)$ and high $P(q)$ condition, C means control condition; PC means probability-conscious group, non-PC means the complement of the PC group, CC means consistent conditional interpretation group, non-CC means the complement of the CC group. MLE means the maximum likelihood estimates of subjective probabilities. Each of them is detailed in the text. See the note of Table 4 on G^2 and P.

information in the task. Using or not using particular information can be seen as a kind of strategy. The non-PC group can be seen to adopting the strategy of ignoring probability information.

The card selection proportions for the PC group in Condition H–H are $R(p) = .53$, $R(\overline{p}) = .67$, $R(q) = .47$, and $R(\overline{q}) = .80$, as shown in Table 5. This result (i.e., high $\neg q$ and $\neg p$) coincides with the QODS model's prediction, but not with that of the ODS model. The estimated values of $P(p)$ and $P(q)$ were .59 and .60, respectively. These are high values and are consistent with the probabilistic information given in the task.

In Condition L–H, the card selection pattern of the PC group does not coincide with the predictions of either the QODS model or the ODS model. The estimated values were $P(p) = .44$, $P(q) = .44$. This result may be explained as follows. In terms of probability, a logical rule "r: $p \rightarrow q$" is represented by a joint probability distribution shown in Table 1 (left), which is a case of Table 2 (left) with $u = P(q \mid r_0, p) = 1$ as described earlier, and $v = P(p \mid r, q)$ can take any arbitrary value, $0 \leq v \leq 1$. As a special case, if $v = 1$ then the rule satisfies the biconditional relationship "r: $p \leftrightarrow q$". So, from a probabilistic viewpoint, a *nearness* to a biconditional relationship can be defined. As v increases, "$p \rightarrow q$" can be regarded as coming closer to "$p \leftrightarrow q$". In Condition L–H, where $P(p)$ = LOW and $P(q)$ = HIGH, v should be low according to standard probability theory. As a logical consequence the relationship differs from a biconditional—that is, it violates the biconditionality assumption. In other words, in Condition L–H, there was an incompatibility between the biconditionality assumption and the probabilistic information given in the task. Such a conflict may explain why the expected results were not obtained in this condition.

Evaluated values of probabilities. Because the PC group's results are very similar to those of all participants for the evaluated values of $P(p)$, $P(q)$, and $P(r)$, the following analyses were only performed for the whole group of participants.

The mean evaluated values of $P(p)$ and $P(q)$ were .28 and .60 in Condition L–H, .70 and .68 in Condition H–H, and .42 and .40 in Condition C, respectively. Only in Condition L–H was the difference between $P(p)$ and $P(q)$ significant, $t(43) = 6.15$, $p < .01$. The difference in the $P(q)$ value between Conditions L–H and H–H was not statistically significant, $t(89) = 1.67$, $p = .10$. Thus, at least superficially, participants properly comprehended the probabilistic information that was given in the task. However, the evaluated values disagreed with the estimated values of these probabilities in Condition L–H. This may be caused by the difference between a participant's superficial, or explicit, understanding of given information and the way information is used by the processes responsible for card selection.

The rating of $P(r)$ varied according to the conditions: .45, .66, and .52, in Conditions L–H, H–H, and C, respectively, $F(2, 129) = 5.96$, $p < .01$. Multiple comparisons by Tukey's HSD procedure indicated that only the difference between L–H and H–H was statistically significant ($p < .05$). This result is explained by the biconditionality assumption. People may not only have a tendency to interpret a conditional biconditionally, they may also believe it to be the norm. Under this assumption, the believability of the information that describes a *nonbiconditional* relationship will be lower than the believability of the information that describes a biconditional one. Only Condition L–H was a nonbiconditional situation, so the low value of $P(r)$ in Condition L–H can be explained.

Statistical independence of card selections. Some studies have argued for statistical associations between card selections (Evans, 1977; Oaksford & Chater, 1994; Pollard, 1985). Because the QODS model separately predicts the probability of each card to be selected, if the model is correct, the statistical independence of card selections should be observed. Each of the four cards can be considered to be a binary categorical variable: to select or not to select. Therefore, the results of an experiment are presented as a four-dimensional $2 \times 2 \times 2 \times 2$ contingency table. Note that it can be misleading, or at least insufficient, to analyse only the marginal tables of a multi-way table (see, e.g., Bishop, Fienberg, & Holland, 1975), which is exactly what was done in the previously mentioned studies. The experimental data for each combination of condition, task, and group were analysed using the log-linear model of independence, forbidding any interaction terms. For the PC group, the model fitted the data for Condition H–H, $G^2(11, N = 15) = 15.96, p = .14$, and for Condition C, $G^2(11, N = 8) = 11.10, p = .43$, whereas an inferior fit was shown in Condition L–H, $G^2(11, N = 21) = 31.8, p < .01$. On the other hand, for the non-PC group, the model fitted none of the data well, $G^2(11, N = 24) = 25.01, p < .01$, $G^2(11, N = 32) = 24.33, p = .01$, $G^2(11, N = 33) = 26.82, p < .01$, in Conditions L–H, H–H, and C, respectively. These results indicate that the participants in the PC group performed as predicted in conditions other than L–H, but none of the non-PC group's results can be explained by the QODS model. In Experiment 2, the possible factors responsible for the results in the L–H condition were investigated.

EXPERIMENT 2

In the L–H condition in Experiment 1, the main reason for the lack of an effect of probabilistic information was the conflict between the participants' tendency to a biconditional interpretation, $P(p) \simeq P(q)$, and the probabilistic information provided in the task, $P(p) \ll P(q)$. This poses the question: Can probabilities be manipulated in a way that overcomes the tendency to a biconditional interpretation? The biconditional interpretation may be suppressed by emphasizing the probabilistic information in the task. To aid participants' comprehension of a nonbiconditional relationship, it might be necessary not only to present the information that $P(p)$ = LOW, $P(q)$ = HIGH, but also to cite some supplementary information, for example, that $v = P(p \mid r, q) \ll 1$. This experiment tested this hypothesis.

Clarifying the meaning (i.e., nonbiconditionality) of the rule by textual manipulations may increase $\neg q$ card selections, though not dramatically (Bracewell, 1974, and Mosconi & D'Urso, 1974, described in Green, 1995a, b; Hoch & Tschirgi, 1985; Margolis, 1987; Platt & Griggs, 1993). Explaining the rule in natural language is related to emphasizing probabilistic information in understanding the nonbiconditional relationship. The effectiveness of explaining that the converse of the rule is not true when presented with probabilistic information was therefore also investigated.

When the biconditionality assumption is broken, different strategies for solving the task may become salient. Therefore, some new indexes were introduced to probe participants' strategies, and the relation between their strategies and task performance was also investigated.

Method

Participants and design

The participants were 85 undergraduates at Ritsumeikan University. Three were later excluded from the analyses, one who had prior experience of the selection task and two who provided invalid responses in the answer sheet. The students were randomly assigned to one of two groups: 40 of the participants were allocated to the *rule explication condition* (Condition E, described later); and 42 participants were allocated to a *control condition* (Condition C).

As in Experiment 1, the experiment was conducted in groups and the tasks were printed in a booklet. The instructions stated that the task was to be performed according to page sequence and never in the reverse order.

Materials and procedure

The tasks were similar to those of Experiment 1. They differed only in the following points. First, the evaluation task for the probabilities $P(p)$, $P(q)$, and $P(r)$ was performed at the start. This change was intended to call the participants' attention more to the probabilistic information and to promote comprehension of the nonbiconditional relationship suggested in the rule.

Second, an additional questionnaire on the evaluation of the probability of p given q when the hypothesis is true, that is, $v = P(p|r, q)$, was introduced: "Now suppose you are drawing some trash cards from a new deck of Hanafuda cards. What is the ratio of red cards to black cards when the rule Taniyama claims is true?" The ordinary selection task was conducted after this questionnaire.

Third, in addition to Questions 1 and 2 as in Experiment 1, participants were given the following 5-point scale question:

3. Did you consider that the converse of Taniyama's rule was not always obeyed (i.e., the back of a trash card is not always red)?

The intention here was to test participants' sensitivity to the nonbiconditionality of the rule.

In the explication condition (Condition E), the following sentence was inserted in the instruction directly after the rule (with no additional sentence in Condition C).

> What must be noticed is that the converse of the rule is not always obeyed. That is to say, the back of a trash card is not necessarily red, and a card with a reverse side that is black may be a trash card or a scoring card.

Results and discussion

The effect of rule explication. In Condition E, the proportions of card selection were $R(p) = .88$, $R(\overline{p}) = .38$, $R(q) = .58$, and $R(\overline{q}) = .55$; in Condition C, they were $R(p) = .69$, $R(\overline{p}) = .21$, $R(q) = .62$, and $R(\overline{q}) = .57$. Except for the p card, none of the differences between corresponding proportions of the two conditions were statistically significant, $\chi^2(1, N = 82) = 4.07, 2.56, 0.17$, and 0.04, respectively ($p = .04$ for the p card, and ps $> .05$ for the others). Although it could be expected that the selection frequency of the $\neg q$ card would rise in Condition E, this was not observed. These results indicate that verbal explication of a conditional (i.e., nonbiconditional) relationship only had a weak effect on changing card selection frequency.

Original task. In Condition E, the proportions of cards selected were $R(p) = .88$, $R(\overline{p}) = .28$, $R(q) = .68$, and $R(\overline{q}) = .28$; and in Condition C, they were $R(p) = .98$, $R(\overline{p}) = .29$, $R(q) = .83$, and $R(\overline{q}) = .21$. No significant differences was detected between the conditions, $\chi^2(1, N = 82) = 3.09$, 0.01, 2.79, and 0.41, respectively (ps > .05). As a whole, the proportions were $R(p) = .93$, $R(\overline{p}) = .28$, $R(q) = .76$, and $R(\overline{q}) = .24$. In Experiment 1, they were $R(p) = .88$, $R(\overline{p}) = .39$, $R(q) = .69$, and $R(\overline{q}) = .37$. The differences of corresponding proportions between experiments were not significant for all cards, $\chi^2(2, N = 214) = 1.27$, 2.51, 1.10, and 3.75, respectively (ps > .05). These results corroborate the homogeneity of participants between the conditions and between experiments.

Indexes on strategies and grouping participants. As was explained in the Results and Discussion section of Experiment 1, the value of v, which participants evaluated in the probability evaluation task, can be regarded as an indicator of the similarity of a conditional to the biconditional. Therefore, a *biconditional interpretation index*, BI, for each participant was defined as his or her evaluation of v, which reflects the subjective estimate that the converse of the rule is true.

From Bayes' theorem and the QODS model's assumption,

$$P(q) = P(q|r) = \frac{P(p|r)}{P(p|r,q)} = \frac{P(p)}{P(p|r,q)}$$

If a participant was perfectly Bayesian, his or her evaluated subjective probabilities would satisfy this relationship. Because all these probabilities, that is, $P(p)$, $P(q)$, and $v = P(p \mid r, q)$, were collected in the probability evaluation task, we can examine whether or not such an ideal relation is maintained for each participant. We define a *probabilistic deviation index*, DI, as

$$\left| P(q) - \frac{P(p)}{v} \right|$$

which indicates how well the participant maintains consistency with ordinary probability theory.

Here, let us consider the participants who satisfied low BI and low DI. These participants can be considered to have little tendency towards a biconditional interpretation and to be probabilistically consistent.[7] Such a group of participants may employ a common probabilistic strategy. The high/low values of the indexes were defined in relation to the medians (BI: 20; DI: 40). The group of participants who had a low BI and a low DI were labelled the *consistent conditional interpretation* (CC) group.[8] PI, introduced in Experiment 1, can be looked on as an index that is based on a participant's self-knowledge (i.e., explicit cognition). On the other hand, DI is an index that indicates a participant's consistency in probabilistic calculation and is based on implicit cognition. There appeared to be no dependency between high/low PI ($\geq 4/\leq 3$) and high/low DI, $\chi^2(1, N = 82) = 0.157$, $p = .69$.

[7]In the present experiment, if the probabilistic information given in the task is interpreted precisely in a manner based on ordinary probability theory, the relationship between p and q should be nonbiconditional.

[8]Participants with a value of $P(p)$ higher than .50 and those with a $P(q)$ lower than .50 were also excluded from the CC group because these values suggest an inadequate comprehension of probability.

The number of CC participants differed significantly between conditions: 16/40 (Condition E) vs. 8/42 (Condition C), $\chi^2(1, N = 82) = 4.35, p < .05$. This suggests that the explication by natural language triggered a consistent interpretation of conditionals for some participants.

Evaluated values of probabilities. The mean ratings of $P(p)$ and $P(q)$ were .18 and .67, respectively. As in Experiment 1, the difference was significant, $t(81) = 12.2$, $p < .01$. However, although the probabilistic information given in the task was reflected in the mean rating, they do not coincide with the estimated values. The instructions stated that $P(q)$ = HIGH, whereas the estimates indicated *moderate* values (.51 for the CC group, which is not so high as the evaluated value, .67). This result is discussed further in the General Discussion section.

Card selection frequencies. The proportions of cards selected in each group are shown in Table 5. Across all participants the proportion of $\neg q$ card selections was relatively high (.56) compared with the result of the original task (.24), but it was lower than that for the q card (.60). Although this result differed from the results of Experiment 1, it did not coincide with the QODS model's predictions.

However, the CC group showed a significantly higher frequency of $\neg q$ card selections in the probabilistic task than did participants in the non-CC group: $R(\overline{q})$ = .83 and .45, respectively, $\chi^2(1, N = 82) = 10.22$ and $p < .01$. The estimates for the CC group were $P(p) = .43$ and $P(q) = .51$, which reflect (to a certain extent) the effect of the probabilistic information given in the task.

It is also apparent that in the original task (which does not contain probabilistic information) no clear differences in card selections were observed between the CC group and the non-CC group: $\chi^2(1, N = 82) = 0.05, 0.02, 3.16$, and 1.10, for the p, $\neg p$, q, and $\neg q$ cards, respectively (ps > .05). In addition, the estimates of the original task for these groups were similar: $P(p)$ = .25 and $P(q) = .27$ for the CC group; $P(p) = .24$ and $P(q) = .25$ for the non-CC group. From these observations, it follows that in a selection task with an effective probabilistic manipulation, the CC group participants demonstrated the selection tendency that was predicted by the model. On the other hand, in a task without any probabilistic information, they made a similar selection as that made by the non-CC group, based on the rarity assumption (Oaksford & Chater, 1994).

Raising $P(p)$ and $P(\overline{q})$ will increase the chance of giving the logically correct answer, namely, selecting only the p and $\neg q$ cards. On the probabilistic task, the CC group demonstrated a higher proportion correct (.33) than did the non-CC group (.16), whereas on the original task, this difference disappeared (.00 and .07, respectively).

In this respect, the results for the PC group closely resemble those for the CC group. For the PC group, the estimates in the probabilistic task were $P(p) = .37$ and $P(q) = .44$, and those for the original task were $P(p) = .26$ and $P(q) = .28$. The proportion correct was .30 in the probabilistic task and .07 in the original task. Therefore, the QODS model successfully predicts the card selection tendency of participants who are consistent from a probabilistic point of view or who are conscious of probability.

Statistical independence of card selection. To examine the statistical independence of card selections, selection patterns were analysed using the log-linear model. The independence model fitted both the CC, $G^2(11, N = 27) = 14.4, p = .42$, and the PC group data, $G^2(11, N = 27) = 14.4, p = .21$, but fitted neither the non-CC, $G^2(11, N = 58) = 35.4, p < .01$, nor the non-PC group data, $G^2(11, N = 55) = 24.2, p = .01$. These results confirm that the performance of the CC and PC group participants is well predicted by the QODS model.

GENERAL DISCUSSION

A new computational model of the selection task with an indicative conditional rule ("if p then q") was presented in this article. The model was a quantitative revision of the ODS model that permitted estimation of the subjective probabilities of the antecedent (p) and the consequent (q) of the rule from experimental data. Meta-analyses using the model revealed that, generally, $P(p)$ and $P(q)$ were approximately equal in the standard abstract task, which was called the biconditionality assumption. Experiments 1 and 2 were run to test the superiority of the new model and to examine the validity of new methods to sort participants by their strategies.

The results of the experiments showed that the probability information given in a task is not sufficient to change responses in the selection task for all participants. In Experiment 2, the CC group's performance was as the model predicts, and the opposite to the results in the L–H condition of Experiment 1. This reveals that participants' performance can be affected by how well they comprehend the task situation, which can be controlled by emphasizing probabilistic information as in Experiment 2 (on the topic of encoding of the problem information, see Green, 1995b; Platt & Griggs, 1993, 1995). Comprehension may also be affected by familiarity with the task scenario or by how believable it is.

In a recent paper, Feeney and Handley (2000) introduced another indirect method to affect participants' comprehension of the probabilistic situation of the task. In addition to a standard conditional rule "if p_1 then q", they presented a second rule "if p_2 then q". In their experiments, because the probabilistic information itself was not manipulated at all, it is unclear whether, and how, the relations between $P(p_1)$ and $P(q)$ changed when presented with or without the second rule. Their results, however, can still be explained by the QODS model, as opposed to the ODS model. Typically, two cases can be considered: In the two-rule condition, compared with the one-rule case, (1) the participants lowered $P(p_1)$, and (2) they heightened $P(q)$. In either case, selection tendencies of the q card and $\neg p$ card are decreased monotonically as shown in Figure 2, agreeing with Experiments 2 and 3 in Feeney and Handley (2000).

There are some problems with providing probabilistic information in the selection task. One is how to present the information. For example, Gigerenzer and Hoffrage (1995) discussed the superiority of frequency formats. Careful attention needs to be paid not only to what information is presented, but also to how it should be presented, because information has an effect only if it can be comprehended by participants. Another problem is that of how the information is understood. Kahneman and Tversky (1979) have pointed out the difference between a stated probability and a perceived subjective probability. As discussed in the section Meta-analysis Using the Model, their π-function straightforwardly indicates the general tendency to superadditiveness that coincides with the results of the estimation of subjective probabilities.

There has been much debate on the effects of prior beliefs on the selection task (Chater & Oaksford, 1999; Evans & Over, 19996a, b; Klauer, 1999; Oaksford & Chater, 1998; Oaksford et al., 1999; Pollard & Evans, 1983). Evans and Over (1996a) criticized the ODS model for not being "general enough" because it assumes that the alternative hypothesis is only that p and q are independent (p. 90). This excludes innumerable other alternative relationships. However, as we have seen in the section Modelling Alternative Hypotheses, modelling the competition among hypotheses did not provide very meaningful results. A model should be preferred and usually sufficient if it is the simplest one that involves just a dependence and independence hypothesis. However, in contrast to the model's prediction that prior beliefs have almost no effect in changing behaviour, a number of studies have shown the effects of prior beliefs, especially on $\neg q$ selection in the selection task (Fiedler & Hertel, 1994; Love & Kessler, 1995; Pollard & Evans, 1981, 1983; Van Duyne, 1976). Prior belief, $P(r)$, has been manipulated in various ways in a number of studies. Fiedler and Hertel (1994) controlled it by creating the *suspicion* that it was false, by mentioning rule violations (e.g., "there are rumours that negate the rule") in the instruction. Oaksford et al. (1999) evaluated $P(r)$ using rating tasks. However, in some studies (e.g., Pollard & Evans, 1981, 1983), it was confused with $P(q \mid r, p)$ (Chater & Oaksford, 1999) or $P(q \mid p)$. In these studies, low prior belief meant low $P(q \mid p)$, or high $P(\overline{q} \mid p)$, which was associated with the rule "if p then *not* q" ($p \rightarrow \neg q$). It may be reasonable to examine the results of such studies in the framework of the three-hypothesis model introduced in the section Modelling Alternative Hypotheses. Apart from the discussion of the distinction between $P(q \mid p)$ and $P(r)$, however, both prior belief and the index of conditional dependency have so weak an effect on changing the card selection tendencies, as we have seen, that the model cannot account for the increase of $\neg q$ selection in the high $P(\overline{q} \mid p)$ condition.

Though there have been no clear and definitive theories of these results up to now, one of the most likely accounts in the framework of the QODS model is as follows. The participants so strongly believed the rule $p \rightarrow \neg q$ that their task became to verify it, as opposed to the experimenter's intention, which was to test $p \rightarrow q$. In addition, as Oaksford and Chater (1996) pointed out, in the training phase of the Pollard and Evans (1983) experiment, as participants' attention was directed to the probability of the occurrence of qs on the other side of p cards, $P(q \mid p)$, information on $P(p)$ and $P(q)$ may have been ignored. Consequently they adopted the default rarity values. As shown in Experiment 1, probabilistic information without appropriate emphasis hardly affects performance, especially when it deviates from rarity. Given the rule $p \rightarrow \neg q$ with $P(p)$ = LOW and $P(\overline{q})$ = LOW, the QODS models predicts high selections of the p and $\neg q$ cards.

In Fiedler and Hertel (1994), it seems that $P(r)$ was properly manipulated, so the above account cannot apply to their results. However, Oaksford et al. (1999) did not find the effects of prior beliefs with systematic manipulations of $P(p)$ and $P(q)$. Considered along with the results of some other experiments (e.g., Legrenzi, 1971; Mosconi & D'Urso, 1974, described in Green, 1995a), it can be concluded that the effects of $P(r)$ are not robust. Generally, the results of a selection task experiment should be regarded as a mixture of various strategies, as argued earlier. One of the most probable ways of accounting for the diverse results may be that strategy usage may be influenced by information about suspicion.

Oaksford et al. (1999) properly distinguished $u = P(q \mid r, p)$ and $P(r)$ as, in our terms, the *unexceptionality* and *believability* of the rule. It is important to distinguish these when analysing or modelling human reasoning involving conditional sentences. However, it may also be true that, usually, people can hardly distinguish these two subjective probabilities on their own. Only in regard to the following two special cases may one be able to discriminate these probabilities: a belief in a complete dependence rule with a little suspicion versus a belief in an incomplete dependence rule with no suspicion. For example, it is possible for a sceptical person to doubt the completely dependent rule "if there is no unbalanced force acting on a particle, then it remains at rest or continues to move in a straight line with a uniform velocity". In this case, $u = 1$ and $P(r) \neq 1$. On the other hand, one can adopt an absolute belief in a rule, "Birds fly", that permit some exceptions. Here, $u \neq 1$ and $P(r) = 1$. However, when one has an incomplete belief (i.e., with any suspicion) in an incomplete dependent rule (e.g., "Polychlorinated biphenyl affects the reproductive health of human beings"), it would be very hard to discriminate between u and $P(r)$.

As a method for making this distinction clearer, in Experiments 1 and 2, participants rated $P(r)$ as the believability of the rule claimed by a specified person (who is described as an assistant in the instructions). In evaluation tasks, it is important to ensure that participants do not confuse u and $P(r)$. At the same time, however, it is important to keep in mind the reliability or the validity of participants' explicit (meta-)cognition of their own subjective probabilities. It is obvious that participants do not perform explicit arithmetical operations on subjective probabilities, but they do perform implicit computations. The subjective probabilities used in implicit cognitive processes probably cannot always be verbalized accurately. In Experiments 1 and 2, the evaluated subjective probabilities indicated extreme values, but the estimates were *moderate*. These results show a dissociation between explicit and implicit cognition. Therefore, we should not place too much trust in the values provided by participants.

Another rational model of the selection task proposed by Klauer (1999) addresses an important issue related to prior belief and exceptionality. In his model, a *risk function* is involved. It defines the risk of the decision procedure (an expected loss caused by possible errors in hypothesis testing, i.e., the Type I and II errors in standard statistics) and the cost of experimentation. When $u \neq 1$, a "p and $\neg q$" case cannot refute the rule, and it *can* be considered to be an "exception". For example, someone who has a (false) belief that "if I bring this talisman I will always perform well" may hardly change his belief after experiencing some negative cases that he "brought the talisman and performed badly". The extent to which one will tolerate exceptions may be a matter of costs and benefits (i.e., risk), and of *epistemic utility* (Evans & Over, 1996a, b). In the selection task, with semantically rich material, this problem may become significant.

The moderate results for the parameter estimates (i.e., the superadditiveness of subjective probabilities) in the meta-analyses and Experiments 1 and 2 can be understood from the standpoint of adaptive rationality. Moderateness relates to *conservativeness* (cf., Edwards, 1968) or insensitivity to the given information. The differences in participants' estimated values of $P(p)$ and $P(q)$ (i.e., $|P(p) - P(q)|$) were not very high. These results are consistent with those for probability estimation in Kirby's (1994) experiment, described in the section Meta-analysis Using the Model, where it was shown that $P(p) \simeq P(q)$. This suggests that participants find it difficult to enlarge $|P(p) - P(q)|$. Put another way, it is difficult to lower $v = P(p \mid r, q)$, or to decrease the *nearness* to the biconditional relationship of the conditional rule. This indicates a tendency to resist *computing* the most optimal behaviour in an unusual situation where the rarity assumption is violated. If, in most cases of daily life, p and q are rare and they have a biconditional relationship, then moderateness is no longer a bias, but may be an adaptive strategy, and thus rational in the sense of Anderson (1990). Indeed, according to the analysis of Klayman and Ha (1987), people's natural *positive test heuristic* is effective when $P(p)$ and $P(q)$ are both low and similar in value. If one computes the most adaptive strategy every time one confronts a hypothesis, total costs will add up. So, insensitivity to probabilistic information could be adaptive in an environment where rarity is the norm. In the *complex* environment in which we live, where no one can always predict all changes, adaptivity cannot be defined *a priori*. Therefore, the unpredictability of the environment may be a cause of diversification of strategies, and it might be in the nature of living things. Thus, the present study has explored, at one strategy level, the sensitivity of people to probabilistic information, taking the discussion forward from the work of Oaksford et al. (1997).

In summary, it was found that the following components are important to comprehend performance on the abstract selection task: (1) the rarity assumption, (2) the biconditionality assumption, and (3) moderateness, or insensitivity to changes in the environment. The QODS model together with these three key concepts, each of which is an aspect of human adaptivity to the environment, can explain a great deal of the experimental data on the abstract selection task.

REFERENCES

Almor, A., & Sloman, S.A. (1996). Is deontic reasoning special? *Psychological Review*, *103*, 374–380.

Anderson, J.R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Beattie, J., & Baron, J. (1988). Confirmation and matching biases in hypothesis testing. *The Quarterly Journal of Experimental Psychology*, *40A*, 269–297.

Berger, J.O. (1985). *Statistical decision theory and Bayesian analysis* (2nd ed.). New York: Springer-Verlag.

Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: The MIT Press.

Bracewell, R.J. (1974, April). *Interpretation factors in the four card selection task*. Paper presented at the selection task conference, Trento, Italy.

Chater, N., & Oaksford, M. (1999). Information gain and decision-theoretic approaches to data selection: Response to Klauer (1999). *Psychological Review*, *106*, 223–227.

Cheng, P.W., & Novick, L.R. (1991). Causes versus enabling conditions. *Cognition*, *40*, 83–120.

Cheng, P.W., & Novick, L.R. (1992). Covariation in natural causal induction. *Psychological Review*, *99*, 365–382.

Chrostowski, J.J., & Griggs, R.A. (1985). The effects of problem content, instructions, and verbalization procedure on Wason's selection task. *Current Psychological Research and Reviews*, *4*, 99–107.

Cover, T.M., & Thomas, J.A. (1991). *Elements of information theory*. New York: John Wiley & Sons.

Earman, J. (1992). *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge, MA: The MIT Press.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: John Wiley & Sons.

Evans, J.St.B.T. (1977). Toward a statistical theory of reasoning. *The Quarterly Journal of Experimental Psychology*, *29*, 621–635.

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. (1996a). *Rationality and reasoning*. Hove, UK: Psychology Press.

Evans, J.St.B.T., & Over, D.E. (1996b). Rationality in the selection task: Epistemic utility versus uncertainty reduction. *Psychological Review*, *103*, 356–363.

Feeney, A., & Handley, S.J. (2000). The suppression of q card evidence: Evidence for deductive inference in Wason's selection task. *The Quarterly Journal of Experimental Psychology*, *53A*, 1224–1242.

Fiedler, K., & Hertel, G. (1994). Content-related schemata versus verbal-framing effects in deductive reasoning. *Social Cognition*, *12*, 129–147.

Geis, M.L., & Zwicky, A.M. (1971). On invited inferences. *Linguistic Inquiry*, *2*, 561–566.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684–704.

Global Optimization 3.0 [Computer software]. (1998). Napervill, IL: Loehle Enterprises.

Green, D.W. (1995a). The abstract selection task: Thesis, antithesis, and synthesis. In S.E. Newstead & J.St.B.T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honour of Peter Wason* (pp. 173–188). Hove, UK: Lawrence Erlbaum Associates Ltd.

Green, D.W. (1995b). Externalization, counter-examples, and the abstract selection task. *The Quarterly Journal of Experimental Psychology*, *48A*, 424–446.

Green, D.W., & Over, D.E. (1997). Causal inference, contingency tables and the selection task. *Current Psychology of Cognition*, *16*, 459–487.

Green, D.W., & Over, D.E. (1998). Reaching a decision: A reply to Oaksford. *Thinking and Reasoning*, *4*, 187–192.

Green, D.W., Over, D.E., & Pyne, R.A. (1997). Probability and choice in the selection task. *Thinking and Reasoning*, *3*, 209–235.

Griggs, R.A. (1984). Memory cueing and instructional effects on Wason's selection task. *Current Psychological Research and Reviews*, *3*, 3–10.

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

Hattori, M. (1999). The effects of probabilistic information in Wason's selection task: An analysis of strategy based on the ODS model. In *Proceedings of the Second International Conference on Cognitive Science and the 16th Annual Meeting of the Japanese Cognitive Science Society Joint Conference (ICCS/JCSS99)* (pp. 623–626).

Hoch, S.J., & Tschirgi, J.E. (1985). Logical knowledge and cue redundancy in deductive reasoning. *Memory & Cognition*, *13*, 453–462.

Howson, C., & Urbach, P. (1993). *Scientific reasoning: The Bayesian approach* (2nd ed.). Peru, Illinois: Open Court. (Original work published 1989.)

Japan Publications Inc. (1970). *Hanafuda: The flower card game.* Tokyo: Nichibo shupan-sha.

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

Johnson-Laird, P.N., & Wason, P.C. (1970). Insight into a logical relation. *The Quarterly Journal of Experimental Psychology*, *22*, 49–61.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kirby, K.N. (1994). Probabilities and utilities of fictional outcomes in Wason's four-card selection task. *Cognition*, *51*, 1–28.

Klauer, K.C. (1999). On the normative justification for information gain in Wason's selection task. *Psychological Review*, *106*, 215–222.

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

Laming, D. (1996). On the analysis of irrational data selection: A critique of Oaksford and Chater (1994). *Psychological Review*, *103*, 364–373.

Legrenzi, P. (1971). Discovery as a means to understanding. *The Quarterly Journal of Experimental Psychology*, *23*, 417–422.

Love, R.E., & Kessler, C.M. (1995). Focusing on Wason's selection task: Content and instruction effects. *Thinking and Reasoning*, *1*, 153–182.

Manktelow, K.I., & Evans, J.St.B.T. (1979). Facilitation of reasoning by realism: Effect or non-effect? *British Journal of Psychology*, *70*, 477–488.

Margolis, H. (1987). *Patterns, thinking, and cognition: A theory of judgment*. Chicago, IL: The University of Chicago Press.

Mathematica 4.0 [Computer software]. (1999). Champaign, IL: Wolfram Research, Inc.

McKenzie, C.R.M., & Mikkelsen, L.A. (2000). The psychological side of Hempel's paradox of confirmation. *Psychonomic Bulletin and Review*, *7*, 360–366.

Mosconi, G., & D'Urso, V. (1974, April). *The selection task from the standpoint of the theory of double code*. Paper presented at the conference on the selection task, Trento, Italy.

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

Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review*, *103*, 381–391.

Oaksford, M., & Chater, N. (1998). A revised rational analysis of the selection task: Exceptions and sequential sampling. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 372–398). London: Oxford University Press.

Oaksford, M., Chater, N., & Grainger, B. (1999). Probabilistic effects in data selection. *Thinking and Reasoning*, *5*, 193–243.

Oaksford, M., Chater, N., Grainger, B., & Larkin, J. (1997). Optimal data selection in the reduced array selection task (RAST). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 441–458.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 883–899.

Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 835–854.

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.

Over, D.E., & Evans, J.St.B.T. (1994). Hits and misses: Kirby on the selection task. *Cognition*, *52*, 235–243.

Over, D.E., & Jessop, A.L. (1998). Rational analysis of causal conditionals and the selection task. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 399–414). London: Oxford University Press.

Platt, R.D., & Griggs, R.A. (1993). Facilitation in the abstract selection task: The effects of attentional and instructional factors. *The Quarterly Journal of Experimental Psychology*, *46A*, 591–613.

Platt, R.D., & Griggs, R.A. (1995). Facilitation and matching bias in the abstract selection task. *Thinking and Reasoning*, *1*, 55–70.

Pollard, P. (1985). Nonindependence of selections on the Wason selection task. *Bulletin of the Psychonomic Society*, *23*, 317–320.

Pollard, P., & Evans, J.St.B.T. (1981). The effects of prior beliefs in reasoning: An associational interpretation. *British Journal of Psychology*, *72*, 73–81.

Pollard, P., & Evans, J.St.B.T. (1983). The effect of experimentally contrived experience on reasoning performance. *Psychological Research*, *45*, 287–301.

Read, T.R.C., & Cressie, N.A.C. (1988). *Goodness-of-fit statistics for discrete multivariate data*. New York: Springer-Verlag.

Stanovich, K.E. (1999). *Who is rational? Studies of individual differences in reasoning*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

Taplin, J.E. (1971). Reasoning with conditional sentences. *Journal of Verbal Learning and Verbal Behavior*, *10*, 219–225.

Valentine, E.R. (1985). The effect of instructions on performance in the Wason selection task. *Current Psychological Research and Reviews*, *4*, 214–223.

Van Duyne, P.C. (1976). Necessity and contingency in reasoning. *Acta Psychologica*, *40*, 85–101.

Wason, P.C. (1966). Reasoning. In B.M. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Harmondsworth: Penguin Books.

Wason, P.C. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology*, *20*, 273–281.

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

Yachanin, S.A. (1986). Facilitation in Wason's selection task: Content and instructions. *Current Psychological Research and Reviews*, *5*, 20–29.

Original manuscript received 16 March 2001
Accepted revision received 19 July 2001

APPENDIX

Expected information gain expressed as a function of $P(p)$, $P(q)$, and $P(r)$

Generally speaking, $I(X;\, Y\,|\,z)$, the *average mutual information* measure between x and y conditioned by z, is defined by the following equations:

$$
\begin{aligned}
I(X;Y|z) &= H(X|z) - H(X|Y,z)\\
&= \sum_{XY} P(x_i, y_j|z)\log\frac{P(x_i|y_j,z)}{P(x_i|z)}
\end{aligned}
$$

Here, log x means $\log_2 x$, and the value of 0 log 0 is assumed to be 0.

Let $P = \{p_i\} = \{p, \overline{p}\}, Q = \{q_j\} = \{q, \overline{q}\}, R = \{r_k\} = \{r, \overline{r}\}$, and x, y, and z represent $P(p)$, $P(q)$, and $P(r)$, respectively. Then, $I(Q;\, R\,|\,p)$—the expectation of uncertainty reduction by turning over the p (e.g., vowel) card—which is abbreviated to $I(p)$, can be calculated by the following equations:

$$
\begin{aligned}
I(p) &\triangleq I(Q;R|p)\\
&= \sum_{QR} P(q_j, r_k|p)\log\frac{P(q_j,r_k|p)}{P(q_j|p)P(r_k|p)}\\
&= \frac{P(p,q,r)}{P(p)}\log\frac{P(p,q,r)P(r)}{P(p,q)P(p,r)} + \frac{P(p,q,\overline{r})}{P(p)}\log\frac{P(p,q,\overline{r})P(\overline{r})}{P(p,q)P(p,\overline{r})}\\
&\quad + \frac{P(p,\overline{q},r)}{P(p)}\log\frac{P(p,\overline{q},r)P(r)}{P(p,q)P(p,r)} + \frac{P(p,\overline{q},\overline{r})}{P(p)}\log\frac{P(p,\overline{q},\overline{r})P(\overline{r})}{P(p,\overline{q})P(p,\overline{r})}\\
&= y\overline{z}\log y - \overline{y}\overline{z}\log\overline{z} - (y\overline{z} + z)\log(y\overline{z} + z).
\end{aligned}
$$

Here, $\overline{x}$, $\overline{y}$, and $\overline{z}$ indicate $(1 - x)$, $(1 - y)$, and $(1 - z)$, respectively. $I(\overline{p})$, $I(q)$, and $I(\overline{q})$ are also calculated similarly. They can be described as a function of x, y, and z as follows:

$$
\begin{aligned}
I(\overline{p}) &= \overline{z}\log\overline{x} + y\overline{z}\log y + \frac{(y-x)z}{\overline{x}}\log(y-x)\\
&\quad - \frac{\overline{y}(\overline{x}\overline{z} + z)}{\overline{x}}\log(\overline{x}\overline{z} + z) - \frac{x\overline{y}\overline{z} + (y-x)}{\overline{x}}\log[x\overline{y}\overline{z} + (y-x)]\\
I(q) &= \overline{z}\log y + \overline{x}\overline{z}\log\overline{x} + \frac{(y-x)z}{y}\log(y-x)\\
&\quad - \frac{x(y\overline{z} + z)}{y}\log(y\overline{z} + z) - \frac{x\overline{y}\overline{z} + (y-x)}{y}\log[x\overline{y}\overline{z} + (y-x)]\\
I(\overline{q}) &= \overline{x}\overline{z}\log\overline{x} - x\overline{z}\log\overline{z} - (\overline{x}\overline{z} + z)\log(\overline{x}\overline{z} + z).
\end{aligned}
$$