THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2003, *56A* (6), 1079–1088

Deciding between accounts of the selection task: A reply to Oaksford (2002)

Aidan Feeney
University of Durham, UK
Simon Handley
University of Plymouth, UK
Robert W. Kentridge
University of Durham, UK

In this paper we report on our attempts to fit the optimal data selection (ODS) model (Oaksford & Chater, 1994; Oaksford, Chater, & Larkin, 2000) to the selection task data reported in Feeney and Handley (2000) and Handley, Feeney, and Harper (2002). Although Oaksford (2002b) reports good fits to the data described in Feeney and Handley (2000), the model does not adequately capture the data described in Handley et al. (2002). Furthermore, across all six of the experiments modelled here, the ODS model does not predict participants' behaviour at the level of selection rates for individual cards. Finally, when people's probability estimates are used in the modelling exercise, the model adequately captures only 1 out of 18 conditions described in Handley et al. We discuss the implications of these results for models of the selection task and claim that they support deductive, rather than probabilistic, accounts of the task.

How best to think about how people should reason has, in the last 10 years, been the subject of much debate (see Anderson, 1990; Evans, 2002; Oaksford & Chater, 2001). This is an important argument for cognitive scientists in general, as it goes to the heart of the rationality question, and for experimental psychologists in particular, as how we think people should reason drives the experiments we design to see how people do reason. The question boils down to whether reasoning tasks require people to be good decision makers or good logicians and nowhere has the debate been more marked, or the issues more clearly delineated, than in the literature on Wason's selection task (Wason, 1966). In the selection task people are presented with a conditional rule and four cards. They are told that the rule specifies what occurs on each

Requests for reprints should be sent to Aidan Feeney, Applied Psychology, University of Durham, Queen's Campus, Thornaby, Stockton-on-Tees, TS17 6BH, UK. Email: aidan.feeney@durham.ac.uk or Simon J. Handley, Centre for Thinking and Language, Department of Psychology, University of Plymouth, Drake's Circus, Plymouth PL4 8AA, UK. Email: shandley@plymouth.ac.uk

We are very grateful to Mike Oaksford for sending us a copy of his Mathematica program. Robert W. Kentridge is The University of Durham Sir Derman Christopherson Foundation Fellow.

http://www.tandf.co.uk/journals/pp/02724987.html DOI:10.1080/02724980245000034

side of the cards and are asked which card(s) they need to turn over in order to test whether the rule is true or false. For example, the rule might be "If there is an A on one side of the card then there is a three on the other", whilst the visible sides of the four cards might contain an "A", an "L", a "3", and a "7". The traditional view of this task is that people should turn over those cards that are capable of disconfirming the rule. That is, they should choose to turn over the "A" card, as a "not-3" on the other side would disconfirm the rule, and the "7" card, as an "A" on the other side would also disconfirm the rule. Very few people make this pattern of selections, and this has led to the claim that logical reasoning does not determine the cards that people elect to turn over (e.g., see Evans, 1989).

Oaksford and Chater (1994) suggested that a more appropriate normative analysis of the task was one where participants sought to maximize the expected information to be gained from selecting each card. That is, Wason's task should be thought of as a decision-making problem where the probabilities associated with the antecedent and consequent in the rule determine people's behaviour. Oaksford and Chater presented the optimal data selection (ODS) model of the task, which uses the Shannon–Weiner measure of information and Bayes' theorem to predict what cards people should select on the task if they are attempting to maximize expected information gain. In general, they achieve very good fits with the data. However, there is continuing debate on whether the selection task is best thought of as a decision making or a logical task and on the specifics of the Oaksford and Chater model (see Evans & Over, 1996; Klauer, 1999; Laming, 1996; Poletiek, 2000).

In two papers (Feeney & Handley, 2000; Handley, Feeney, & Harper, 2002) we described the results of six experiments designed to demonstrate that, contrary to the claims of decision-theoretic accounts of Wason's task, people do attempt to deduce the logical consequences of turning over the cards in the selection task. However, Oaksford (2002b) applies the ODS model to the experiments reported by Feeney and Handley (2000) and reports good fits to the data. He argues that Feeney and Handley were incorrect in claiming that their data posed difficulties for Oaksford and Chater's (1994) information gain model and that they should have performed quantitative modelling of the data rather than making qualitative predictions. In our reply to Oaksford we report on our attempts to fit the ODS model to the results of three experiments described in Handley et al. (2002). We report our analysis of the predictions made by the ODS model for each card across all six of the experiments reported in Feeney and Handley and Handley et al. We argue that the results of this modelling exercise do provide us with evidence that discriminates between rival accounts of the task. Finally, we discuss the status of the ODS account and the distinction drawn by Oaksford between qualitative and quantitative predictions. First, however, we briefly describe the nature of the experimental evidence that is at issue.

The data

Feeney and Handley (2000) reported three experiments applying the suppression paradigm (see Byrne, 1989; Rumain, Connell, & Braine, 1983) to Wason's selection task (Wason, 1966). Alongside the rule to be tested, *If p then q*, participants were given a second rule, *If r then q*, which specified an alternative antecedent for the consequent to occur. To borrow Oaksford's (2002b) example, participants might have been asked to test the rule *if the key is turned (p) then the car starts (q)* but were also told that *if the car is hot-wired (r) then the car starts*. We expected that participants would be less likely to select the *not-p* and the *q* cards in two rule conditions because the presence of the alternative antecedent would make them less likely to infer: (1) that *the car has not started (not-q)* given that *the key has not been turned (not-p)*; and (2) that *the key was turned (p)* given that *the car has started (q)*. In Experiments 2 and 3 of our paper we found that the presence of a second rule suppressed the rate at which the *not-p* and *q* cards were selected. Our interpretation of these findings is that people are explicitly considering what might be on the other side of the card when making their selections. We argued that these results supported accounts of selection task performance which have participants deducing the consequences of the various outcomes possible given that particular cards are turned over. We suggested that the results are problematic for the ODS model as our number of rules manipulation is likely to increase the perceived probability of q, Pr(q), and, according to our qualitative understanding of the model, this should result in an increase in *not-q* card selections as well as suppression of *q* card selections. Our reading of the model was that it did not make any predictions about *not-p* card selections as it assumed that this card was always uninformative.

Oaksford (2002b) takes issue with the predictions that we qualitatively induced from the ODS model. He points out that recent formulations of the model (Oaksford, Chater, & Larkin, 2000) allow the *not-p* card to be informative and by fitting the model to our data show that it predicts an increase in *p* card selections and a reduction in *q* and *not-p* card selections. The second and third of these predictions are borne out by our results, and, in general, the model provides excellent fits to our data. However, we are concerned that when treating the data from Feeney and Handley (2000) meta-analytically by condition, there are very few data points. In fact, the model's predictions for each card were based upon only nine data points. Accordingly, we have concerns about Type I errors. In addition, the experimental designs used in Handley et al. (2002) provide more statistical power than those employed in Feeney and Handley.

One of our aims in this paper is to apply the ODS model to the experiments described in Handley et al. (2002), thereby testing how well the model fits data from more powerful studies, as well as enabling a test of the model's meta-analytic predictions based upon a greater number of data points. We did not design the experiments described in Feeney and Handley (2000) as tests of the ODS account specifically (although Experiment 3 is probably the best test of the model in that paper), and accordingly, we did not fit the model to the data. To redress this omission we report below the results of our attempts to fit the model to the data described in Handley et al. Importantly, these experiments were specifically designed as tests of the ODS model.

All three of the experiments described in Handley et al. (2002) employed the same basic methodology as that used in our earlier experiments and produced similar results: Participants received problems containing one or two rules, and strong suppression of *q* card selections was associated in all three experiments with the presentation of a second rule. The suppression of *q* card selections was accompanied by *not-p* card suppression in Experiment 1 and a marginally significant decrease in *p* card selections in Experiment 2. In addition to manipulating number of rules, Handley et al. also tested the ODS model by explicitly manipulating Pr(q). These manipulations were achieved either by explicitly varying the size of the antecedent set (Experiments 1 and 2) or including in the scenario that accompanied the tasks reference to antecedent sets other than that referred to in the second rule. Although participants' estimates of Pr(q) reliably differed as a function of these manipulations, the manipulations produced no effects on card selections. We suggested that these results are problematic for the ODS model. However, they disconfirm only the qualitative predictions that we induced from the theory. Therefore we now report on our attempts to fit the ODS model to the data reported in Handley et al. (2002).

The modelling exercise

We fitted the information gain model to the data from Handley et al. (2002) in almost exactly the same fashion as that described in Oaksford (2002b). Briefly, we found the probability values of p and q—we refer to these as Pr(p) and Pr(q) respectively—that maximized the log likelihood given the information gain model. We did this using Mathematica's FindMinimum function (Wolfram, 1991). Using contour plots of fitness landscape we established that, in every case reported, the fitness functions had single maxima. Unlike Oaksford (2002b), we did not therefore supplement Mathematica's gradient search for local minima with a global optimizer. Before conducting analyses of the Handley et al. data we confirmed that this simplified procedure replicated Oaksford's results for Feeney and Handley's (2000) experiments perfectly (see Table 1). Using these best fit values for Pr(p) and Pr(q) we estimated the expected information gain for each card. Next we used Hattori's (1999) "selection tendency function" to transform expected information gain estimates into response probabilities. Following Oaksford, we used the log-likelihood ratio test statistic G^2 to assess how well the model's predictions fit the predictions of a fully saturated model where the probability of selecting any card was set to the observed probability. As only Pr(p) and Pr(q) were estimated from the data, G^2 was assessed against two degrees of freedom, and, once again following Oaksford, the significance level for rejection was set at 1%.

The results of the model fitting for all six of the experiments contained in Feeney and Handley (2000) and Handley et al. (2002) are shown in Table 1. As participants in Experiment 3 of Handley et al. completed three tasks each, we report the data for each problem content separately by condition. The information gain model is rejected in 3 out of 4 of the two-rule conditions in Experiments 1 and 2 of Handley et al. and in 2 out of the 6 two-rule cases from Experiment 3. In three of the remaining cases from the two-rule conditions of that experiment and in the two-rule high Pr(q) condition of Experiment 1, the predicted values are close to deviating significantly from the observed data. We also followed Oaksford (2002b) in carrying out meta-analytic exploration of our data where condition was treated as the unit of analysis. Although number of rules and alternatives in Experiment 3 of Handley et al. were manipulated between participants, problem content was a within-participants variable. For the purposes of our analysis, therefore, we included the mean values collapsed across problem content for each condition of that experiment. Even allowing for this, as our meta-analysis was carried out on the results of all of the experiments reported in Feeney and Handley and Handley et al., it is based on more than twice as many observations as is the analysis reported by Oaksford.

The meta-analysis revealed a significant difference due to number of rules, $t(17) = 3.38$, $p < .005$, in the best fit estimates of Pr(q) but not in the estimates for P(p), $t(17) = 0.03$, $p > .98$. The predicted rate of *q* card selections also differed significantly due to the number of rules, $t(17) = 5.05$, $p < .001$, as did the observed rate of *q* card selections, $t(17) = 6.42$, $p < .001$.

TABLE 1

Results of the model-fitting exercise for all of the experiments in Feeney and Handley (2000) and Handley, Feeney, and Harper (2002)

			Cards												
			p		*not-p*		*q*		*not-q*		Pr(p)	Pr(q)			
Experi-ment	*Rules*		*Obs.*	*Pred.*	*Obs.*	*Pred.*	*Obs.*	*Pred.*	*Obs.*	*Pred.*	*Pred.*	*Pred.*	*Obs.*	$G^2(2)$	*P values*
F&H 1	One abstract		.92	.88	.20	.17	.60	.55	.32	.27	.28	.34	–	1.22	.545
	One thematic		.80	.83	.08	.20	.60	.56	.36	.29	.31	.36	–	3.52	.172
F&H 2	One implicit negation		1	.92	.15	.13	.75	.65	.30	.17	.19	.24	–	6.37	.041
	One explicit negation		.80	.73	.35	.26	.60	.57	.35	.34	.36	.39	–	1.35	.509
	Two implicit negation		.92	.96	.15	.13	.20	.29	.25	.34	.23	.42	–	2.74	.253
	Two explicit negation		.85	.89	.10	.17	.40	.41	.35	.36	.30	.41	–	1.14	.567
F&H 3	One		.93	.86	.28	.18	.63	.58	.30	.26	.29	.34	–	4.51	.105
	Two single extra antecedent		.93	.95	.05	.13	.33	.33	.30	.31	.23	.39	–	3.46	.178
	Two disjunctive extra antecedent		.95	.94	.18	.14	.38	.38	.28	.29	.24	.37	–	.56	.756
HF&H 1	One		.72	.71	.30	.27	.58	.61	.28	.32	.37	.39	.07	.77	.681
	Two low Pr(q)		.78	.86	.15	.19	.40	.51	.20	.32	.30	.38	.12	9.74	.008
	Two high Pr(q)		.77	.84	.15	.20	.40	.48	.27	.35	.33	.40	.26	6.68	.035
HF&H 2	One		.90	.87	.23	.18	.52	.51	.30	.30	.29	.37	.27	2.32	.314
	Two low Pr(q)		.74	.82	.20	.21	.24	.39	.32	.46	.35	.46	.08	17.27	.001
	Two high Pr(q)		.79	.81	.27	.21	.22	.34	.43	.52	.36	.49	.75	9.89	.008
HF&H 3	One few alternatives	University	1	.99	.17	.10	.37	.36	.13	.14	.10	.22	.40	2.29	.318
		Car	1	.98	.10	.10	.43	.41	.17	.14	.12	.22	.39	1.36	.506
		Phone	.97	.97	.20	.11	.47	.49	.10	.15	.13	.22	.44	3.00	.223
	One many alternatives	University	.93	.95	.10	.13	.43	.44	.23	.24	.22	.33	.72	.38	.829
		Car	.90	.92	.07	.15	.50	.50	.27	.25	.24	.33	.63	2.06	.357
		Phone	.87	.92	.13	.14	.47	.54	.13	.22	.23	.30	.67	3.29	.193
	Two few alternatives	University	.90	.97	.13	.12	.27	.36	.13	.23	.19	.34	.25	5.85	.054
		Car	.87	.93	.17	.14	.30	.41	.17	.29	.25	.37	.31	5.83	.054
		Phone	.87	.96	.03	.13	.23	.33	.20	.29	.22	.39	.35	9.79	.008
	Two many alternatives	University	.97	.99	.03	.09	.10	.12	.17	.19	.08	.49	.73	3.91	.142
		Car	.87	.97	.03	.11	.20	.31	.13	.25	.19	.36	.78	13.32	.001
		Phone	.93	.99	.03	.10	.20	.25	.13	.18	.13	.33	.70	6.44	.040

Note: F&H = Feeney and Handley (2000). HF&H = Handley, Feeney, and Harper (2002).

Interestingly, the model also predicted a significantly lower probability of *not-q* card selection in the one-rule conditions than in the two-rule conditions, $t(17) = 2.26$, $p < .05$. This prediction was not borne out by the data, $t(17) = .41$, $p > .5$. Finally, although the model did not predict significant effects of the number of rules manipulation for the *p* card, $t(17) = 1.22$, $p > .2$, or the *not-p* card, $t(17) = .86$, $p > .4$, the effect on observed selection rates for the *not-p* card is approaching significance, $t(17) = 1.79$, $p < .1$. There was no effect of the number of rules manipulation on the probability of the *p* card being selected, $t(17) = .84$, $p > .4$.

As we carried out pretests for the experiments in Handley et al. on the perceived Pr(q) we were able to examine the relationship between participants' estimates and the best fit values from the model. There was almost no correlation between the two sets of estimates ($r = -.02$). This suggests that participants' probability estimates do not match the predictions of the model. We also carried out the modelling exercise on the results of the experiments from Handley et al. using the best fit values for Pr(p) and participants' estimates of Pr(q). As only Pr(p) was estimated from the data, G^2 was evaluated against one degree of freedom. The results of this additional modelling may be seen in Table 2.

Across the experiments reported by Feeney and Handley (2000) and Handley et al. (2002), the ODS model does not fare particularly well. Although the fits to the data in Feeney and Handley are very good, the model does not predict much of the two-rule data described in Handley et al. where participant numbers were substantially greater. In addition, the predictions that the model makes for each card over the six experiments are not borne out by the data. Whilst the ODS model predicts a lower probability of selecting the *q* card and a higher probability of selecting the *not-q* card in the two-rule conditions, the data contain significant suppression of *q* card selections and marginal suppression of *not-p* card selections. These are exactly the predictions that we have argued a deductive account would make for the suppression paradigm as applied to the selection task. Importantly, both the model predictions and the findings about card selection tendencies reported here are based upon a bigger sample than it was possible for Oaksford (2000b) to use.

The fits between the ODS model and our data are even worse when Pr(q) is taken directly from participants' estimates rather than being estimated from the data. In only 1 of 18 cases do the model's predictions accord satisfactorily with the observed pattern of card selections. It is striking that when Pr(q) predicted by the modelling exercise is replaced by participants' estimates, almost all of the fits are poor regardless of the sign or the size of the difference between predicted and estimated Pr(q). Unsurprisingly, given such poor fits when participants' estimates rather than the model's estimates are used, we found no correlation between the best fit values and participants' estimates for Pr(q).

Qualitative vs. quantitative predictions and the status of the ODS model

Oaksford (2000b) distinguishes between qualitative and quantitative predictions about the results of reasoning experiments. We feel that there is an important logical problem with his distinction. Whilst we cannot but agree that we made qualitative predictions about the results described in Feeney and Handley (2000) and Handley et al. (2002), we do not feel that Oaksford has made quantitative predictions. Instead he has, post hoc, estimated the fit between his model and the data. To do this he has used the data to estimate best fit values

TABLE 2

Results of the model-fitting exercise using participants' estimates of Pr(q) for all three experiments in Handley, Feeney, and Harper (2002)

			Cards									
			p		*not-p*		*q*		*not-q*			
Experiment	*Rules*		*Obs.*	*Pred.*	*Obs.*	*Pred.*	*Obs.*	*Pred.*	*Obs.*	*Pred.*	$G^2(1)$	*P values*
HF&H 1	One		.72	.92	.30	.10	.58	.82	.28	.10	77.13	.0001
	Two low Pr(q)		.78	.97	.15	.10	.40	.62	.2	.11	47.39	.0001
	Two high Pr(q)		.77	.91	.15	.14	.40	.63	.27	.19	26.95	.0001
HF&H 2	One		.90	.92	.23	.14	.52	.60	.30	.19	13.32	.0003
	Two low Pr(q)		.74	.98	.20	.09	.24	.52	.32	.10	143.29	.0001
	Two high Pr(q)		.79	.97	.27	.10	.22	.10	.43	.65	82.65	.0001
HF&H 3	One few alternatives	University	1	.98	.17	.11	.37	.25	.13	.26	7.27	.0070
		Car	1	.97	.10	.12	.43	.29	.17	.28	6.65	.0099
		Phone	.97	.95	.20	.13	.47	.29	.10	.37	16.85	.0001
	One many alternatives	University	.93	.99	.10	.09	.43	.10	.23	.33	29.36	.0001
		Car	.90	.98	.07	.10	.50	.12	.27	.46	35.18	.0001
		Phone	.87	.99	.13	.09	.47	.11	.13	.34	46.93	.0001
	Two few alternatives	University	.90	.98	.13	.10	.27	.38	.13	.16	7.37	.0066
		Car	.87	.95	.17	.13	.30	.45	.17	.22	7.10	.0077
		Phone	.87	.96	.03	.12	.23	.36	.20	.25	10.28	.0013
	Two many alternatives	University	.97	1.0	.03	.09	.10	.09	.17	.20	4.28	.0385
		Car	.87	.99	.03	.09	.20	.09	.13	.28	26.52	.0001
		Phone	.93	1.0	.03	.09	.20	.09	.13	.21	13.23	.0003

Note: HF&H = Handley, Feeney, and Harper (2002). For predicted Pr(p) and Pr(q) associated with each condition, see Table 1.

for Pr(p) and Pr(q). He then uses these values to make "predictions". Importantly, these predictions cannot be made without the data, and hence it is doubtful as to whether they are predictions at all. However, we feel that the model-fitting exercise is very important in that it appears to have confirmed the quantitative basis for our qualitative prediction that the use of the suppression paradigm would discriminate between accounts of the selection task. Across all six experiments modelled here, the ODS account predicts *q* card suppression and an increase in *not-q* card selection. Importantly, and contrary to the claim made by Oaksford concerning the experiments reported in Feeney and Handley (2000), it does not appear to make any prediction about changes in *not-p* card selection when the results for all six experiments are considered. On the other hand, deductive accounts of the task would predict *not-p* and *q* card suppression owing to the presence of a second rule. Therefore, the data would appear to support deductive accounts and to disconfirm the ODS account. This conclusion seems even safer when one considers how poorly the ODS model fits the two-rule data from Handley et al. (2002).

The ODS account of the selection task was originally devised in the spirit of Anderson's (1990) rational analysis of cognition. Accordingly, it may be understood as being about the goals of the computations carried out by people faced with the selection task. The ODS account has been remarkably successful in modelling performance on the selection task, suggesting, as Oaksford and his colleagues have argued, that modal performance on the task is rational when judged against an alternative normative standard to logic. Of course, as is the case with any computational level account, there is no guarantee that the algorithm that implements the computation will always be successful. Given the fact that ODS fits most of our single-rule data—that is, if we overlook the modelling based upon subjective estimates of Pr(q), one possible means of reconciling the account with the findings here would be to claim that the effect that the introduction of a second rule has on selection rates is an issue that is best addressed at the algorithmic rather than the computational level (see, Oaksford, 2002a, for similar claims regarding procedural variations on the conditional arguments task). That is, perhaps the introduction of a second rule represents such a significant procedural variation that the algorithm that implements ODS (whatever it might be) fails to maximize information gain in the two-rule case. We would find any claim of this kind highly implausible, and given Oaksford's (2002b) claim that the effects of a second rule can be readily captured by ODS, this is perhaps a line of argument that few would be comfortable with. Our view on the basis of the modelling presented here is that the ODS account is not the appropriate way of characterizing participants' behaviour whether they are presented with one or with two rules.

The most parsimonious view of participants' behaviour is that the same processes underlie selections in both the single and two-rule cases. This behaviour involves considering what one could deduce about what would be on the other side of any given card assuming that the rule was true. The pragmatics of a conditional, "if p then q" are such that it often invites one to infer that "If q then p" and "if not-p then not-q" also hold (Geis & Zwicky, 1971; Horn, 2000). On a conditional arguments task these invited inferences lead to the affirmation of the consequent and the denial of the antecedent fallacies. On the selection task participants generally only consider the *p* and the *q* cards, because they do not recognize that inferences can be made from implicitly negated cards (see, for example, Evans & Handley, 1999). Hence participants in the single-rule condition select the *p* card, inferring that there will be a q on the other side, and the *q* card, inferring that there will be a p on the other side. This second inference is blocked in the two-rule case leading to a reduction in *q* card selections. Of course some participants will process the implicit negation on the cards and consider the *not-p* card, inferring in the one rule case that there should be a not-q on the other side. Again this inference is blocked in the two-rule case. In simple terms, the presence of a second rule with an alternative antecedent blocks a biconditional interpretation of the rule (Rumain et al., 1983) leading to a reduction in *q* card selections and a reduction in *not-p* card selections for those participants who have successfully processed the negation. Whilst this explanation is descriptive in nature, it could easily be framed in terms of mental models (Johnson-Laird & Byrne, 1991) or relevance theory (Sperber, Cara, & Girotto, 1995), as we have discussed elsewhere (Handley et al., 2002). The key point is that this simple explanation of the effect of a second rule provides a unified explanation of the data from both the selection task and the conditional arguments task.

It is worth pointing out here (as did an anonymous reviewer of an earlier draft of this paper) that being able to decide which norm people's card selections best approximate is of importance in protecting the selection task from recent criticism (e.g., Sperber et al., 1995; Sperber & Girotto, 2002) that it represents "a sunk cost in the history of the psychology of reasoning" (Sperber & Girotto, 2002, p. 289). We have shown that rival accounts of human reasoning make different predictions about people's performance on the selection task, and we have been able to demonstrate that participants' behaviour more closely follows one set of predictions than the other. Although the task may be over-used, it also appears to discriminate between theories of human reasoning.

Finally we wish to emphasize that, although our data are inconsistent with the predictions of the ODS account, we would not wish to dismiss probabilistic approaches wholesale. Indeed there is ample evidence that in many cases human behaviour on reasoning tasks can be shown to respect the probabilistic structure of the environment or of the task (e.g., Feeney, Evans, & Clibbens, 2000; Stevenson & Over, 1995). However, in this paper we have done just as Oaksford (2002b) suggests and checked that our qualitative understanding of the ODS approach corresponds to the quantitative predictions that the model makes. The most important claim that we wish to make is indeed a quantitative one: The optimal data selection model does not fit the data.

REFERENCES

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

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition*, *31,* 61–83.

Evans, J. St. B. T. (1989). *Biases in human reasoning: Causes and consequences*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J. St. B. T. (2002), Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin*, *128*, 978–996.

Evans, J. St. B. T., & Handley, S. J. (1999). The role of negation in conditional inference. *Quarterly Journal of Experimental Psychology*, *52A*, 739–769.

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

Feeney, A., Evans, J. St. B. T., & Clibbens, J. (2000). Background beliefs and evidence interpretation. *Thinking and Reasoning*, *6*, 97–124.

Feeney, A., & Handley, S. J. (2000). The suppression of q card selections: Evidence for deductive inference in Wason's selection task. *Quarterly Journal of Experimental Psychology*, *53A*, 1224–1243.

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

Handley, S. J., Feeney, A., & Harper, C. (2002). Alternative antecedents, probabilities, and the suppression of fallacies in Wason's selection task. *Quarterly Journal of Experimental Psychology*, *55A*, 799–818.

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

Horn, L. R. (2000). From if to iff: Conditional perfection as pragmatic strengthening. *Journal of Pragmatics*, *32*, 289–326.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* Hove, UK: Lawrence Erlbaum Associates Ltd.

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

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

Oaksford, M. (2002a). Contrast classes and matching bias as explanations of the effects of negation on conditional reasoning. *Thinking and Reasoning*, *8*, 135–151.

Oaksford, M. (2002b). Predicting the results of reasoning experiments: Reply to Feeney and Handley (2000). *Quarterly Journal of Experimental Psychology*, *55A*, 793–798.

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. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, *5*, 349–357.

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

Poletiek, F. (2001). *Hypothesis testing behaviour*. Hove, UK: Psychology Press.

Rumain, B., Connell, J., & Braine, M. D. S. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: IF is not the biconditional. *Developmental Psychology*, *19*, 471–481.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, *52*, 3–39.

Sperber, D., & Girotto, V. (2002). Use or misuse of the selection task? Rejoinder to Fiddick, Cosmides, and Tooby. *Cognition*, *85*, 277–290.

Stevenson, R. J., & Over, D. E. (1995). Reasoning from uncertain premises. *Quarterly Journal of Experimental Psychology*, *48A*, 613–43.

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

Wolfram, S. (1991). *Mathematica* [Computer program]. Reading, MA: Addison-Wesley.

Original manuscript received 6 August 2002
Accepted revision received 6 November 2002