THINKING AND REASONING, 1998, *4* (3), 193–230

Cognitive Ability and Variation in Selection Task Performance

Keith E. Stanovich
University of Toronto, Canada

Richard F. West
James Madison University, Virginia, USA

Individual differences in performance on a variety of selection tasks were examined in three studies employing over 800 participants. Nondeontic tasks were solved disproportionately by individuals of higher cognitive ability. In contrast, responses on two deontic tasks that have shown robust performance facilitation—the Drinking-age Problem and the Sears Problem—were unrelated to cognitive ability. Performance on deontic and nondeontic tasks was consistently associated. Individuals in the correct/correct cell of the bivariate performance matrix were over-represented. That is, individuals giving the modal response on a nondeontic task (P and Q) were significantly *less* likely to give the modal response on a deontic task (P and not-Q) than were individuals who made the non-modal P and not-Q selection on nondeontic problems. The implications of the results are discussed within the heuristic-analytic framework of Evans (1996; Evans & Over, 1996) and the optimal data selection model of Oaksford and Chater (1994).

INTRODUCTION

One of the most extensively studied tasks in the psychology of reasoning is Wason's (1966) selection task (sometimes termed the "four-card task"). The participant is shown four cards lying on a table showing two letters and two numbers (A, D, 3, 7). They are told that each card has a number on one side and a letter on the other and that the experimenter has the following rule (of the "if P, then Q" type) in mind with respect to the four cards: "If there is an A on one side

Requests for reprints should be sent to Keith E. Stanovich, Department of Human Development and Applied Psychology, Ontario Institute for Studies in Education, University of Toronto, 252 Bloor St. West, Toronto, Ontario, Canada M5S 1V6. Email: KStanovich@oise.utoronto.ca

This research was supported by a grant from the Social Sciences and Humanities Research Council of Canada to Keith E. Stanovich and a James Madison University Program Faculty Assistance Grant to Richard F. West. The authors thank Walter Sa for assistance in data coding. Jonathan Evans, Richard Griggs, Mike Oaksford, and Richard Platt are thanked for their extensive comments on an earlier version of the manuscript.

then there is a 3 on the other". The participant is then told that he/she must turn over whichever cards are necessary to determine whether the experimenter's rule is true or false (for recent introductions to the vast literature on the selection task, see Evans, Newstead, & Byrne, 1993; Newstead & Evans, 1995). Performance on such abstract versions of the selection task is extremely low (see Beattie & Baron, 1988; Evans, 1989; Griggs & Cox, 1982; Jackson & Griggs, 1988, 1990; Wason & Evans, 1975). Typically, less than 10% of individuals make the correct selections of the A card (P) and 7 card (not-Q). The most common incorrect choices are the A card and the 3 card (P and Q) or the selection of the A card only (P).

Early on in the investigation of the selection task it was thought that the use of a real-life, but arbitrary rule ("Every time I go to Manchester I travel by train") would facilitate performance, but subsequent research has demonstrated that this is not the case (Dominowski, 1995; Evans, 1989, 1995; Griggs & Cox, 1982; Manktelow & Evans, 1979). However, one particular rule introduced by Griggs and Cox (1982) has consistently produced substantially improved performance. When testing the rule "if a person is drinking beer, then the person must be over 19 years of age" and when given the four cards beer, Coke, 22, and 16 to represent P, not P, Q, and not Q, respectively (hereafter termed the Drinking-age Problem), performance is markedly superior to that on the abstract selection task (Dominowski, 1995; Evans, 1989; Evans et al., 1993; Griggs & Cox, 1982, 1983; Pollard & Evans, 1987).

A vigorous debate has ensued over which theory can explain the robust content effects observed with rules of the type exemplified in the Drinking-age Problem. Cosmides (1989; see also Gigerenzer & Hug, 1992) has argued that such rules activate a modular algorithm concerned with social exchange. This algorithm is geared to the generic rule "if you take a benefit then you must pay a cost" and has sensitive procedures for detecting cheaters—those who violate the rule by taking the benefits without paying the cost. In contrast, Cheng and Holyoak (1985; Cheng, Holyoak, Nisbett, & Oliver, 1986) argue that the Drinking-age Problem triggers a pragmatic reasoning schema—a content-bound set of production rules that have been abstracted from experience with similar situations. Performance on the Drinking-age Problem is facilitated because it fits a pre-existing set of production rules for permission schemas (Holyoak & Cheng, 1995a, b); whereas rules with familiar but arbitrary content ("Every time I go to Manchester I travel by train") would not trigger the same set of pre-existing production rules and hence would not lead to facilitation. For the purposes of this paper, the similarities between these theories are more important than their differences (for extensive discussions of the ongoing controversy, see Chater & Oaksford, 1996; Cummins, 1996; Holyoak & Cheng, 1995a, b; Liberman & Klar, 1996; Over & Manktelow, 1995; Sperber, Cara, & Girotto, 1995).

Although quite different in their specifics, both of these theories seem to have the implication that the type of reasoning (or task interpretation) involved when deontic rules are used in selection tasks (reasoning about rules used to guide human behaviour—about what "ought to" or "must" be done, see Manktelow & Over, 1991) is different from the type of reasoning (or task interpretation) involved when abstract rules are employed (see Griggs & Cox, 1993; Manktelow & Over, 1990). Manktelow, Sutherland, and Over (1995, p.201) argue that:

> deontic reasoning can be contrasted with indicative reasoning roughly along the lines of the philosophical distinction between theoretical and practical reasoning: theoretical reasoning aims to infer what is, was, or will be the case; practical reasoning, or deontic reasoning in the present context, aims to infer what one should, may, or must do.

Manktelow et al. (1995) discuss how the focus on the fulfilling of obligations and permissions by actors and enforcers (see Gigerenzer & Hug, 1992; Oaksford & Chater, 1994) is leading to the merging of deontic theories with decision-making perspectives emphasising the maximisation of expected utility (see Kirby, 1994). But such theories lead inevitably to the question of "whether subjects do much indicative or theoretical reasoning as such in selection tasks" (Manktelow et al., 1995, p.204).

This quote from Manktelow et al. (1995) is simply the latest example of a long line of conjectures that the reasoning styles and/or task interpretations triggered by deontic versions of the selection task are fundamentally different from those implicated in abstract versions (see Griggs, 1983; Tweney & Yachanin, 1985). However, other investigators see an underlying continuity in the processing of deontic and nondeontic tasks. Oaksford and Chater (1994) posit that individuals view the abstract selection task as an inductive problem and not a problem in deductive hypothesis testing. They show that under the inductive construal, a Bayesian model of optimal data selection fits individual choices quite well. For the deontic selection task, Oaksford and Chater (1994) developed the Bayesian model in terms of the utilities of the cards rather than their information value. They demonstrated that under certain assumptions, their model could predict the predominant choice of P and Q in the abstract selection task and P and not-Q in the deontic versions of the task. Their view is that in both tasks people are essentially optimal information processors as they perform a task that they construe as inductive and probabilistic rather than deductive. Patterns of performance in *both* abstract and deontic versions are optimal on a rational analysis that emphasises optimisation to the environment rather than conformity to a normative model (Anderson, 1990, 1991; Oaksford & Chater, 1994, 1995). Thus, at the broad level of optimality there is a commonality between both versions of the task; however, Oaksford and Chater (1994) are silent on the issue of individual differences across the two versions of the task—that is, whether we might expect associations between the tendency towards optimal information gain in the indicative version of the task and optimal expected utility assessment in the deontic version.

In his heuristic-analytic framework, Evans (1984, 1989, 1995, 1996) posits that people use the same generic mechanism to solve both tasks. He proposes a two-stage process of heuristic, followed by analytic, processing (the implications of the discussion that follows also hold for the more parallel conception of Evans & Over, 1996). In the first stage, basic stimulus relevance is determined by preconscious heuristic processes. A conscious analytic process only operates later to justify focusing on the cards that have been given attention due to the heuristic relevance judgement ("analytic reasoning, while present, does not alter the choices made and serves only to rationalise or confirm them, Evans, 1995, p.169). Evans (1995, p.169) hypothesises that heuristics deriving from linguistic function cue responses in the abstract selection task, and that in deontic tasks "card choices are still determined by relevance, but relevance is now cued pragmatically and not linguistically." Thus, although the triggering cues differ, choices in both types of task are determined by heuristic processing.

It will be argued here that an examination of individual differences among participants who complete both types of selection task might enable an elaboration of the models of Evans (1995, 1996; Evans & Over, 1996) and Oaksford and Chater (1994). Both models predict that the modal response pattern of individuals completing both tasks should be P and not-Q on deontic tasks and P and Q on nondeontic tasks (or, collectively, the set of responses traditionally considered "incorrect" on this task, [P]; [P, Q, not-Q]; etc.). Beyond this group of modal individuals, there may be processing implications depending on the distribution of participants in the remaining cells (individuals getting both items correct; those getting the deontic problem incorrect and the nondeontic problem correct; etc.). For example, the individuals in the remaining cells may simply represent error variance around the modal response—that is, probabilistic straying from optimal data selection as conceived in the model of Oaksford and Chater (1994). If this is the case, the bivariate distribution of responses should be characterised by statistical independence. Alternatively, some individuals straying from the modal bivariate response might represent a subgroup of individuals who are systematically deviant from the response predicted by the rational model of Oaksford and Chater (and the two-process view of Evans) because they *are* viewing the nondeontic task as a problem of deductive logic and are reasoning analytically to their choices (rather than just rationalising the heuristically determined responses, as in the two-process view). If such a subgroup exists, then we might expect the cell representing correct responding on both deontic and nondeontic tasks to be over-represented. Thus, unlike the view that individuals deviating from the modal response are error variance, the subgroup hypothesis predicts a dependence in responding across the two tasks. The magnitude of the dependence in responding could be viewed as an index of the size of the subgroup of analytic responders. In fact, the Bayesian model of optimal data selection of Oaksford and Chater (1994) might fit even better if the subgroup of analytic responders could be identified and eliminated from the analysis.

Deriving these predictions highlights a point that has been insufficiently explored in the selection task literature—that patterns of individual differences might throw light on the nature of the mental mechanisms utilised in the deontic and nondeontic versions of the selection task (however, see Evans, 1977; Oaksford & Chater, 1994; Pollard, 1985). Indeed, a further prediction immediately comes to light. That is, if there is a subgroup of analytic responders, the correct/correct cell in the bivariate distribution of responses on deontic and nondeontic tasks should not only be overpopulated, but we might expect it to contain the more cognitively able individuals—if we are willing to make the plausible assumption that these individuals would be more likely to employ analytic reasoning to override the response triggered by their heuristic processing and treat the task as a deductive problem. This assumption is supported by the empirically demonstrated connection between cognitive ability and the tendency to engage in analytic processing (Ackerman & Heggestad, 1997; Cacioppo, Petty, Feinstein, & Jarvis, 1996).

In the three studies reported here, we examined whether performance on deontic and nondeontic selection tasks is related and how performance on various selection tasks relates to a standard index of cognitive ability. In Study 1, performance on several versions of a selection task using thematic but nondeontic content was compared with performance on a version of the Drinking-age rule—a selection task problem that has shown robust facilitation effects. In the studies that follow, cognitive ability was operationalised by an academic aptitude measure (the Scholastic Aptitude Test, SAT) that loads highly on psychometric g—that is, general intelligence. Matarazzo (1972) views the SAT primarily as a measure of general intelligence. Carroll (1993) concurs but suggests that the test is weighted towards crystallised intelligence in the context of the psychometric theory of fluid-crystallised intelligence (Horn & Cattell, 1967). Related measures have been linked to neurophysiological and information-processing indicators of efficient cognitive computation (Caryl, 1994; Deary & Stough, 1996; Detterman, 1994; Hunt, 1987; Vernon, 1993).

STUDY 1

Method

Participants

The participants were 349 undergraduate students (137 males and 212 females) recruited through an introductory psychology participant pool at a medium-sized state university. Their mean age was 18.7 years (SD = 1.2).

Cognitive Ability Measure

Scholastic Aptitude Test Scores. Students were asked to indicate their verbal and mathematical SAT scores on a demographics sheet. The mean reported verbal SAT score (SAT-V) of the students was 529 (*SD* = 72), the mean reported mathematical SAT score (SAT-M) was 578 (*SD* = 72), and mean total SAT score was 1107 (*SD* = 108). These reported scores match the averages of this institution (520, 587, and 1107) quite closely (Straughn & Straughn, 1995; all SAT scores were from administrations of the test prior to its recent rescaling). The Scholastic Aptitude Test is a three-hour paper-and-pencil exam used for university admissions testing. The standardised scores on the verbal and mathematical sections are added together to form the total score. In the entire population of test takers throughout the previous two decades, total scores have averaged approximately 950 with a standard deviation of approximately 150 (Willingham, Lewis, Morgan, & Ramist, 1990). Thus the scores of the students matriculating to this institution are roughly one standard deviation above the mean of all of the prospective university students taking the test.

Selection Task Problems

We employed six items—five with nondeontic rules and the Drinking-age Problem. The former were composed of real-life but somewhat arbitrary content. Because performance on nondeontic problems can sometimes be floored (Griggs & Cox, 1982) we used the "whether or not the rule is being violated" form of the instructions rather than the "test whether the rule is true or false", version because the former has sometimes been shown to facilitate performance (although the facilitation for violation instructions is most often observed when there are other facilitating elements in the problem, see Griggs, 1989; Platt & Griggs, 1993b). One problem (termed the Destination Problem) was as follows:

> Each of the tickets below has a destination on one side and a mode of travel on the other side. Here is a rule: "If 'Baltimore' is on one side of the ticket, then 'plane' is on the other side of the ticket." Your task is to decide which tickets you would need to turn over in order to find out whether or not the rule is being violated. (Mark the appropriate letters.)

Destination: **Baltimore**	Destination: **Washington**	Mode of Travel: **Plane**	Mode of Travel: **Train**
a = turn b = not turn	a = turn b = not turn	a = turn b = not turn	a = turn b = not turn

All five problems were accompanied by a graphic choice like the one in this example. The instructions were parallel to those in this example. The order of the four alternatives that represented the choices P, not-P, Q, and not-Q varied from problem to problem. The other four nondeontic rules were: "If it's a 'USA' postcard, then a '20c' stamp is on the other side of the postcard" (termed the Stamp Problem); "Whenever the menu has 'fish'on one side, 'wine' is on the other side" (termed the Menu Problem); "Every coin with 'Madison' on one side has 'library' on the other side" (termed the Coin Problem); and "Papers with an 'A' on one side have the comment 'excellent' on the other side" (termed the Grade Problem).

The Drinking-age Problem was presented like all of the others, using the instructions:

> Each of the boxes below represents a behaviour on one side and an age on the other side. Here is a rule: "If a person is drinking beer then the person must be over 21 years of age." Your task is to decide which boxes you would need to turn over in order to find out whether or not the rule is being violated. (Mark the appropriate letters.)

Age: **22**	Age: **18**	Drink: **Beer**	Drink: **Coke**
a = turn b = not turn	a = turn b = not turn	a = turn b = not turn	a = turn b = not turn

Note that in this version, we presented the Drinking-age Problem in the same way as the nondeontic problems—that is, as a decontextualised (without scenario) rule. Specifically, we did not employ the scenario of a policeman who was checking for violators which has been shown to facilitate performance (Pollard & Evans, 1987). For this reason, it may not be appropriate to call this version of the Drinking-age Problem a deontic task (without the scenario, it might not have been interpreted as such by all participants). We examine a less controversial version (one with scenario) in Study 2.

Procedure

Participants completed the six selection tasks during a single two-hour session in which they also completed some other tasks not part of the present investigation. They were tested in small groups of 2–6 individuals. The six selection problems were completed consecutively and all participants received them in the following order: Destination Problem, Stamp Problem, Menu Problem, Coin Problem, Grade Problem, Drinking-age Problem.

Results

As in other studies using arbitrary or abstract nondeontic content (Beattie & Baron, 1988; Evans et al., 1993; Hoch & Tschirgi, 1985) only a minority of the individuals chose the P and not-Q cards in the nondeontic problems. Specifically, P and not-Q were chosen by 8.0% of the participants on the Destination Problem, 7.5% of the participants on the Stamp Problem, 6.9% of the participants on the

Menu Problem, 5.7% of the participants on Coin Problem, and 7.4% of the participants on the Grade Problem.

The pattern of responses displayed in Study 1 replicated that found in other studies of performance on the selection task (Beattie & Baron, 1988; Jackson & Griggs, 1988; Oaksford & Chater, 1994; Wason & Evans, 1975). For example, on the Destination Problem, 28 participants answered correctly (P, not-Q), 149 participants gave the most common response observed in other studies (P, Q), 43 participants gave another common response (P), 25 participants gave a response that included the correct pair (P, Q, not-Q), 30 participants checked all of the alternatives, and 74 participants gave some other response pattern (many of these combinations included the not-P card). In a meta-analysis of selection task studies using abstract content, Oaksford and Chater (1994) found that the probabilities of choosing the P, Q, not-Q, and not-P cards were .89, .62, .25, and .16, respectively. The probabilities across the five nondeontic problems in Study 1 (.80, .66, .39, .28) were reasonably convergent—the higher proportion of not-Q choices in the present study being perhaps attributable to the use of violation instructions.

As expected, performance on the Drinking-age Problem was substantially higher than that on the five nondeontic problems. The correct choice of the P and not-Q cards was made by 31.5% of the sample. This proportion is lower than that obtained in previous studies that have used the Drinking-age rule (Griggs & Cox, 1982; Pollard & Evans, 1987) because those studies included a scenario prior to the rule in which individuals are to imagine themselves in the role of a policeman checking for violators. As Study 2 will demonstrate, the inclusion of the policeman scenario boosts performance substantially above the 31.5% correct response rate in the present study. Nevertheless, even the version used in Study 1 resulted in over four times as many correct choices as on the nondeontic problems.

Associations with Cognitive Ability

What are the cognitive characteristics of the individuals who make the correct response on the different versions of the selection task? The data displayed in Table 1 relate to this issue. There, the mean total SAT scores of the participants responding correctly and incorrectly to each of the selection task problems are displayed. There were significant differences in SAT scores between those responding correctly and incorrectly on each of the problems. However, the differences were substantially larger for the nondeontic problems than for the Drinking-age Problem (where the difference barely reached significance, $P = .0485$). The magnitude of the SAT difference in the former ranged from 53 points (Destination Problem) to 83 points (Stamp Problem), whereas the difference on the Drinking-age Problem was only 25 points. In terms of effect sizes, the effects in the nondeontic problems ranged from .502 to .795 (Cohen's *d*). Rosenthal and Rosnow (1991, p.446) classify an effect size of .50 as "moderate" and one of .80

as "large". Thus, the effect sizes on the nondeontic problems ranged from moderate to large. In contrast, the effect size on the Drinking-age Problem (.229) is closer to the value that Rosenthal and Rosnow (1991) classify as "small" (.20).

Of course, some of the SAT difference observed on the Drinking-age Problem is due to the fact that a proportion of the correct Drinking-age responders are the high SAT individuals who *also* answered nondeontic problems correctly. Table 2 displays the mean SAT scores of Drinking-age Problem solvers and nonsolvers for only those individuals who answered that particular nondeontic problem *incorrectly*. For example, 321 individuals answered the Destination Problem incorrectly. Of those, 92 answered the Drinking-age Problem correctly and 229 answered it incorrectly. However, as Table 2 indicates, the mean SAT score of

TABLE 1
Mean SAT Total Scores of Participants Who Gave the Correct and Incorrect Responses to the Six Selection Task Problems in Study 1

	Incorrect	*Correct*	*t (347)*	*Effect Size*[a]
Destination Problem	1103 (321)	1156 (28)	2.54**	.502
Stamp Problem	1101 (323)	1184 (26)	3.89***	.795
Menu Problem	1102 (325)	1172 (24)	3.12***	.662
Coin Problem	1103 (329)	1165 (20)	2.49**	.575
Grade Problem	1102 (323)	1168 (26)	3.03***	.619
Drinking-age Problem	1099 (239)	1124 (110)	1.98*	.229

Figures in parentheses are the number of participants giving the correct and incorrect responses.
* = $P < .05$, ** = $P < .025$, *** = $P < .01$, all two-tailed
[a]Cohen's *d*

TABLE 2
Mean SAT Total Scores as a Function of Drinking-age problem performance for Those Participants Who Gave the Incorrect Responses on Various Nondeontic Problems in Study 1

	Drinking-age Problem		
	Incorrect	*Correct*	*t value*
Destination Problem incorrect	1100 (229)	1108 (92)	0.59
Stamp Problem incorrect	1098 (233)	1107 (90)	0.63
Menu Problem incorrect	1098 (232)	1111 (93)	1.00
Coin Problem incorrect	1099 (234)	1113 (95)	1.04
Grade Problem incorrect	1100 (234)	1108 (89)	0.65

Figures in parentheses are the number of participants giving the correct and incorrect responses on the Drinking-age Problem.

those who solved the Drinking-age Problem (1108) was not significantly different from those responding incorrectly on the Drinking-age Problem (1100). This pattern was apparent when the incorrect responders on each of the other four nondeontic problems were examined. In no case did the mean SAT scores of those responding correctly on the Drinking-age Problem exceed those responding incorrectly.

Participants' predominant responses on the five nondeontic problems were classified according the following criteria. Response patterns on each of the five problems were classified into one of six categories: (1) Correct; (2) P, Q; (3) P, Q, not-Q; (4) P; (5) All; (6) Other. If an individual's responses were in the same category for at least three of the five problems they were assigned that category as the predominant response. If a participant did not have at least three responses in one of the categories, then their predominant response pattern was classified as Mixed. The mean SAT scores of the participants as a function of their predominant response on the five nondeontic problems is displayed in Table 3. The 22 participants having the correct response as their dominant response had the highest SAT scores (1185). The next highest SAT scores (mean = 1159) were obtained by participants having no predominant response—the Mixed group. The scores of this group were high because it was composed in part of individuals giving two (but not three) correct responses, two P responses, or two All responses. As the Table indicates, participants giving Correct, All, or P responses tended to have high SAT scores. The most common predominant category was the P, Q response (150 individuals) and the SAT scores of participants in this category tended to be relatively low (1104). The lowest SAT scores were obtained by participants having an Other response pattern as their predominant response (1055). This occurred because individuals in the Other category often included the not-P card in their choices and those who chose not-P tended to have low SAT scores.

TABLE 3
Mean SAT Total Scores as a Function of the Predominant Response Given on the Five Nondeontic Problems and as a Function of the Response Pattern on the Drinking-age Problem in Study 1

Response Pattern	*Five Nondeontic Problems*	*Drinking-age Problem*
Correct	1185 (22)	1124 (110)
All	1151 (24)	1134 (27)
P	1136 (38)	1126 (47)
P, Q, not-Q	1116 (14)	1099 (18)
P, Q	1104 (150)	1099 (64)
Other	1055 (85)	1072 (83)
Mixed	1159 (16)	

Number of participants in parentheses.

The right side of Table 3 displays the mean SAT scores of individuals giving the same set of responses to the Drinking-age Problem. The pattern of individual differences here was somewhat different from that displayed on the nondeontic problems. The mean SAT scores of the participants giving the P and not-Q response on the Drinking-age Problem were exceeded (although not significantly so) by those of two other groups: those responding by turning all of the cards and those responding by turning only the P card. However, one commonality between the nondeontic problems and the Drinking-age Problem was that participants giving the P and Q response (the matching response, see Evans, 1972, 1995, 1996; Evans & Lynch, 1973) and a response in the Other category tended to have low SAT scores.

Associations Among Responses to Different Selection Rules

Not surprisingly, given the differing percentages of correct responders across items, vastly more participants responded correctly on the Drinking-age Problem and incorrectly to nondeontic problems than responded correctly to a nondeontic version and then got the Drinking-age Problem wrong. The first contingency table displayed in Table 4 illustrates the asymmetry. In this comparison, it can be seen that 92 individuals responded correctly to the Drinking-age Problem but incorrectly to the Destination Problem, whereas only 10 individuals responded correctly to the Destination Problem but incorrectly to the Drinking-age Problem. Performance on the two problems was significantly associated [$\chi^2(1) = 15.14$, $p < .001$]. Although 64.3% of the participants getting the Destination Problem correct also got the Drinking-age Problem correct, only 28.7% getting the Destination Problem incorrect got the Drinking-age Problem correct. Another theoretically important way of viewing the association is to note that individuals making the modal selection on the Drinking-age Problem (P and not-Q) were more likely to make a particular *non*-modal selection on the nondeontic problem (again, P and not-Q). As the discrepancies between the observed and expected cell frequencies indicate, there was an excess of individuals in the incorrect/incorrect and correct/correct cells. The number of individuals in the latter was double the number expected based on a model of statistical independence between responses on the two tasks. Additionally, the SAT scores of these individuals were higher than those of the individuals in the other three cells.

The next contingency table displayed in Table 4 indicates that the pattern of relationships was exactly the same when performance on the Stamp Problem and the Drinking-age Problem was examined. There was a significant association between the two and the direction of the association was such that the incorrect/incorrect and correct/correct cells were over-represented. Similarly, the SAT scores of the individuals in the correct/correct cell were substantially higher than those of the individuals in the other three cells.

Associations between performance on pairs of nondeontic tasks were quite high (phi coefficients ranging from .599 to .760).

TABLE 4
Performance Relationships Among Several of the Problems in Study 1

	Drinking -age Problem Incorrect	*Drinking -age Problem Correct*
Destination Problem		
Incorrect	229 (219.8) [SAT = 1100]	92 (101.2) [SAT = 1108]
Correct	10 (19.2) [SAT = 1071]	18 (8.8) [SAT = 1203]

phi coefficient .208 $\chi^2(1) = 15.14$, $P < .001$

	Drinking-age Problem Incorrect	*Drinking-age Problem Correct*
Stamp Problem		
Incorrect	233 (221.2) [SAT = 1098]	90 (101.8) [SAT = 1107]
Correct	6 (17.8) [SAT = 1132]	20 (8.2) [SAT = 1200]

phi coefficient .277 $\chi^2(1) = 26.83$, $p < .001$

Figures in parentheses are the expected frequencies in that cell based on statistical independence, and figures in brackets are the mean SAT scores.

STUDY 2

The results of Study 1 indicated that the relation between correct responses and cognitive ability was much stronger for the nondeontic versions of the selection task than for the Drinking-age Problem. In fact, unless participants answered a nondeontic version of the selection task correctly, those answering the Drinking-age Problem were no likelier at all to be higher in cognitive ability than those not answering the Drinking-age Problem correctly. Additionally, when the relationships between performance on a nondeontic problem and performance on the Drinking-age Problem was examined, there was an excess of participants in the correct/correct cell. As conjectured in the introduction, this pattern could arise because a small group of participants do not view the nondeontic task as a problem of probabilistic hypothesis testing but instead view it as a deductive reasoning problems (i.e. as the instructions direct). Supporting this conjecture is the fact that the SAT scores of the individuals in the correct/correct cell were substantially higher than those of the individuals in the other three cells—a finding that might be expected if it takes additional cognitive capacity to override heuristic processing.

In Study 2 we test a truly deontic version of the Drinking-age Problem—one that includes a scenario that ensures a deontic interpretation. In Study 2 we also extend these analyses to three selection task rules that have been the subject of much controversy and investigation: a nondeontic purely abstract rule without violation instructions; a social exchange deontic problem without prior familiarity introduced by Cosmides (1989); and a "remedial" selection problem introduced by Margolis (1987; see Griggs, 1989). The first is probably the most investigated type of problem in the selection task literature. The second is typically used to investigate social contract facilitation while controlling for prior familiarity with the counterexample to the rule (see Cosmides, 1989; Gigerenzer & Hug, 1992). The Margolis (1987) problem is an interesting facilitator of performance in that it does so without social exchange considerations. Collectively, the five problems examined in Study 2 span the range of problem types that have been widely investigated in the selection task literature.

Method

Participants

The participants were 294 undergraduate students (108 males and 186 females) recruited through an introductory psychology participant pool at a medium-sized state university. Their mean age was 19.0 years (*SD* = 1.3). The mean reported verbal SAT score (SAT-V) of the students was 525 (*SD* = 71), the mean reported mathematical SAT score (SAT-M) was 580 (*SD* = 78), and mean total SAT score was 1105 (*SD* = 111). These reported scores match the averages of this institution (520, 587, and 1107) quite closely (Straughn & Straughn, 1995; all SAT scores were from administrations of the test prior to its recent rescaling)

Selection Tasks

Abstract Problem. An abstract problem with "true/false", rather than "violated", instructions was employed. The specific problem and instructions were adapted from Platt and Griggs (1993b, p.596) and were as follows:

> Each of the boxes below represents a card lying on a table. Each one of the cards has a letter on one side and a number on the other side. Here is a rule: If a card has a vowel on its letter side, then it has an even number on its number side. As you can see, two of the cards are letter-side up, and two of the cards are number-side up. Your task is to decide which card or cards must be turned over in order to find out whether the rule is true or false. Indicate which cards must be turned over.

K	A	8	5
a = turn b = not turn	a = turn b = not turn	a = turn b = not turn	a = turn b = not turn

All of the subsequent problems were accompanied by a graphic choice like the one in this example.

Destination Problem. The Destination Problem with the violation instructions from Study 1 was included in Study 2 with one small modification. To make the instructions parallel to those of the Abstract Problem the phrase "As you can see, two of the cards are destination-side up, and two of the cards are mode of travel-side up" was added.

Margolis Problem. This problem was introduced by Margolis (1987) and studied systematically by Griggs (1989). The instructions and rule (see Margolis, 1987, p.307; Griggs, 1989, p.520) are:

> Four cards have been picked from a mixed pack (some with red backs and some with blue backs). The person who chose the cards was told to obey the following rule: "Pick any four cards except that if a card has a red back it must be at least a 6". You see the following cards lying on the table, with two face down and two face up. Indicate each card that must be turned over in order to be sure whether it violates the rule.

7 of Hearts	5 of Clubs	Blue Back	Red Back
a = turn b = not turn	a = turn b = not turn	a = turn b = not turn	a = turn b = not turn

This version of the selection task contains numerous cues that facilitate performance (see Griggs, 1989, and Platt & Griggs, 1993b, for an extensive discussion). First, the consequent does not refer to one of the four cards, thus reducing matching bias (Evans, 1972; Evans & Lynch, 1973). This version also reduces the likelihood of a reversible reading of the if/then rule (if at least a 6 then red). More importantly, according to Margolis (1987), the picking of the cards from a deck encourages a correct "closed" reading of the task (that the choice is from a limited set of possibilities) rather than an incorrect "open" reading (that one is being asked to choose *how* to search—for example, that red refers to a *category* that one might choose to search exhaustively; see also Beattie & Baron, 1988, Nickerson, 1996). The "violated" instructions are also used in the Margolis Problem. Although these instructions do not uniformly produce facilitation, they do so in problems that already produce some facilitation (Griggs, 1984; Yachanin, 1986). Finally, the rule contains the deontic term "must" (see Manktelow & Over, 1990) which may help to trigger a deontic conditional interpretation of some problems (see Platt & Griggs, 1993a). However, the problem does not have other features that are thought to cue a social contract interpretation (Cosmides, 1989; Gigerenzer & Hug, 1992; Platt & Griggs, 1993a). Specifically, it has no features emphasising cost/benefit social exchange trade-offs (Cosmides, 1989; Platt & Griggs, 1993a), nor does it emphasise the concept of cheater detection (Gigerenzer & Hug, 1992; Platt & Griggs, 1993a). Thus, although it contains many linguistic and pragmatic cues that serve to avoid misinterpretation, it does not contain the features thought to be essential to cueing a social contract interpretation (Cosmides, 1989; Platt & Griggs, 1993a).

Cosmides Problem. The Kaluame problem introduced by Cosmides (1989) was designed to have precisely the features that social contracts require: an emphasis on the cost/benefit logic of social exchange and the ramifications of the failure to catch a cheater. The problem was designed, however, to be culturally unfamiliar to the participant. In its original version, the Kaluame problem is quite long. Platt and Griggs (1993a) have carried out an extensive series of studies examining the effects of the presence or absence of numerous features of the problem. We employed a version of the problem (termed here the Cosmides Problem) quite similar to the negatives present, must present, cost–benefit absent, cheating perspective present version described in their Study 3. It is a version on which they found 50% of their sample to give P and not-Q responses. This version emphasises cheater detection and uses the deontic term "must" but it does not emphasise the benefits and costs of taking the action without paying the cost that were in Cosmides' original version. The problem is as follows:

> You are a Kaluame, a member of a Polynesian culture found only on Maku Island in the Pacific. The Kaluame have many strict laws which must be enforced, and the elders have entrusted you with enforcing them. The elders have made laws governing what people eat. One of these laws is that if a man eats cassava root, then he must have a tattoo on his face. The cards below have information about four young Kaluame men sitting in a temporary camp; there are no elders around. A tray filled with cassava root and molo nuts has just been left them. Each card represents one man. One side of a card tells which food a man is eating, and the other side of the card tells whether or not a man has a facial tattoo. Your job is to catch men who break the law. If any get past you, you and your family will be disgraced. Indicate only those cards you definitely need to turn over to see if any of these Kaluame men are breaking the law.

eats molo nuts	tattoo	no tattoo	eats cassava root
a = turn b = not turn	a = turn b = not turn	a = turn b = not turn	a = turn b = not turn

Drinking-age Problem. This version of the Drinking-age Problem was a more elaborate version than that employed in Study 1. In the previous version, even though the term violated was used, there was no scenario employed that emphasised the detection of transgressors of the rule—as there was in the original Griggs and Cox (1982) version. Pollard and Evans (1987) found that without the scenario (but also without "must" in the rule and "violated" in the instructions) only 21.4% of their participants responded with P and not-Q. The proportion in Study 1 was higher than this (31.5%). However, even with a "must" rule and "violated" instructions, this solution rate was substantially below the 72.5% in the Griggs and Cox (1982) condition and the 71.4% in the Pollard and Evans (1987) condition where the scenario was used—thus replicating the Pollard and Evans finding that the scenario is crucial. In Study 2 we used a scenario version of the Drinking-age Problem employed by Klaczynski and Laipple (1993) and phrased as follows:

> Imagine that you are a police officer on duty, walking through a local bar. It is your job to ensure that the drinking laws are in effect in this bar. When you see a person engaging in certain activities, the laws specify that certain conditions must first be met. One such law is "If a person is drinking beer then the person must be over 21 years of age." Each of the boxes below represents a card lying on a table. There are two pieces of information about a person on each card. Whether or not the person is drinking beer is on one side of the card and the person's age is on the other side. For two of the people, you can see their age, but you cannot see what they are drinking. For the other two people, you can see what they are drinking, but you cannot see their age. Your task is to decide whether or not this law is being broken in the bar. Circle the card or cards you would definitely need to turn over to decide whether or not the law is being broken. You may select any or all of the cards.

Age: **22**	Age: **18**	Drink: **Beer**	Drink: **Coke**
a = turn b = not turn	a = turn b = not turn	a = turn b = not turn	a = turn b = not turn

Procedure

Participants completed the five selection tasks during a single two-hour session in which they also completed some other tasks not part of the present investigation. Unlike Study 1, they did not complete the problems consecutively but had other tasks interspersed between the selection problems. They were tested in small groups of 2–6 individuals. The five selection problems were presented in one of two orders. In order A (nondeontic first) the presentation sequence was: Abstract Problem, Destination Problem, Margolis Problem, Drinking-age Problem, Cosmides Problem. In order B (deontic first) the presentation sequence was: Cosmides Problem, Drinking-age Problem, Margolis Problem, Destination Problem, Abstract Problem. Order A was completed by 150 participants and order B by 144 participants.

Results

The solution rate on the Abstract Problem (11.6%) was actually higher than that on the Destination Problem (8.2%) indicating that, consistent with previous research, using thematic problems rather than abstract problems does not facilitate performance as long as the content is coupled with a nondeontic rule (Griggs, 1983; Evans, 1989; Evans et al., 1993; Manktelow & Evans, 1979). The Margolis Problem was answered correctly by 31.6% of the sample, indicating a substantial facilitation over the abstract version. The Cosmides Problem was answered correctly by 50.3% of the sample, replicating the substantial facilitation that has repeatedly been found for variants of this problem (Cosmides, 1989; Gigerenzer & Hug, 1992; Platt & Griggs, 1993a). This solution rate is almost identical to the 50.0% rate observed in the most comparable condition of Platt and Griggs'(1993a) Study 4 (cost/benefit absent, "must" present, negatives present, cheating perspective present). The Drinking-age Problem was answered correctly by 85.7% of the sample, substantially higher than the 31.5% who solved the version in Study 1 which was presented without the policeman scenario. This finding reinforces the conclusion of Pollard and Evans (1987) that the policeman scenario must be presented along with the drinking content in order to derive the full facilitation provided by this rule. The 85.7% solution rate in the present study is in the range of (indeed somewhat higher than) the solution rates of 72.5% observed by Griggs and Cox (1982) and 71.4% observed by Pollard and Evans (1987).

There were no significant effects of order on the solution rate for the Abstract and the Cosmides Problems. The solution rate was somewhat higher on the Destination Problem when the deontic tasks were presented first [11.8% versus 4.7%; $\chi^2(1) = 4.99$, $P < .05$]). A significant order effect was observed on the Drinking-age Problem. When it appeared as the second problem (subsequent to the Cosmides Problem) in order B, its solution rate was 95.8%, but when it appeared as the fourth problem (subsequent to the two nondeontic problems and the Margolis Problem) its solution rate was only 76.0% [$\chi^2(1) = 23.60$, $P < .001$]. The latter solution rate is still quite high and it is just as high as that obtained by Griggs and Cox (1982) and by Pollard and Evans (1987) in studies of responses to a Drinking-age problem when not preceded by any other problems. Finally, there was a significant order effect on solution rates for the Margolis Problem

[χ^2 (1)= 28.96, $P < .001$]. Performance was better in Order B where the Drinking-age and Cosmides Problems were presented prior to the Margolis Problem. There, the solution rate was 46.5%—compared to a solution rate of 17.3% in order A where the Margolis Problem was presented before the participant had seen the Drinking-age or Cosmides Problems. Attempting to solve two deontic problems first might have encouraged a deontic interpretation of the Margolis Problem. The deontic/nondeontic status of this problem may be ambiguous (it uses the deontic "must" but does not emphasise cost/benefit social exchanges or cheater detection) and the ambiguity might serve to maximise the effects of priming from previous problems.

Associations with Cognitive Ability

The data displayed in Table 5 present the mean total SAT scores of the participants responding correctly and incorrectly to each of the selection task problems. The results regarding cognitive ability did not interact with order of presentation, so the remaining analyses are reported collapsed across order. There were significant differences in SAT scores between those responding correctly and incorrectly on all of the problems except the Drinking-age Problem. For the other four, effect sizes (Cohen's *d*) were moderate to large—ranging from .410 for the Margolis Problem to .859 for the Destination Problem. Collectively, the SAT difference was larger on the two clearly nondeontic problems than on the Margolis (an ambiguous problem) and Cosmides (deontic) problems. As in Study 1, the Drinking-age Problem displayed the smallest SAT difference. In fact, in Study 2 the difference for this problem was slightly in the opposite direction.

Table 6 presents the frequency of the major response patterns as a function of problem type. The mean SAT score of the group displaying each pattern is indicated. Respondents giving the P and not-Q response had the highest SAT

TABLE 5
Mean SAT Total Scores of Participants Who Gave the Correct and Incorrect Responses to the Five Selection Task Problems in Study 2

	Incorrect	*Correct*	*t (292)*	*Effect Size*[a]
Abstract Problem	1098 (260)	1159 (34)	3.05***	.558
Destination Problem	1097 (270)	1190 (24)	4.02***	.859
Margolis Problem	1091 (201)	1136 (93)	3.26***	.410
Cosmides Problem	1080 (146)	1130 (148)	4.01***	.469
Drinking-age Problem	1110 (42)	1104 (252)	–0.30	.050

Figures in parentheses are the number of participants giving the correct and incorrect responses.
*** = $P < .01$, all two-tailed
[a]Cohen's *d*

TABLE 6
Mean SAT Total Scores as a Function of Response Type on the Five Selection Problems of Study 2

	Abstract	*Destination*	*Margolis*	*Cosmides*	*Drinking-age*
Correct	1159 (34)	1190 (24)	1136 (93)	1130 (148)	1104 (252)
P	1153 (40)	1150 (38)	1109 (63)	1120 (28)	1148 (12)
P, Q, not-Q	1076 (12)	1084 (14)	1104 (13)	1085 (17)	1093 (8)
All	1096 (29)	1101 (21)	1063 (12)	1089 (13)	1086 (9)
P, Q	1091 (140)	1095 (144)	1096 (69)	1045 (35)	1036 (5)
Other	1076 (39)	1070 (53)	1062 (44)	1076 (53)	1143 (8)

Number of participants in parentheses.

scores on every problem except the Drinking-age Problem. Interestingly, participants responding by choosing the P-card only had consistently high SAT scores (a pattern also observed in Study 1, see Table 3). They had the second highest SAT scores on all of the problems except the Drinking-age Problem—on which they actually had the highest mean. The matching response (P and Q) was generally associated with low SAT scores. The mean scores of this group never exceeded 1100 and they decreased as the problems got easier—the mean score was 1045 on the Cosmides Problem and was as low as 1036 on the Drinking-age Problem. Individuals who persisted in giving the matching response even on the easiest problems were disproportionately individuals of lower cognitive ability.

The response of choosing P, Q, not-Q and that of choosing all of the cards are sometimes considered to be superior incorrect responses. The former is often viewed as indicating “partial insight” (Evans, 1977; Wason, 1969), whereas the latter is consistent with a biconditional interpretation of the rule. However, participants giving these two response were no higher in cognitive ability than those giving the matching response (P and Q).

Associations Among Responses to Different Selection Rules

The first three contingency tables displayed in Table 7 indicate that the relationships displayed in Study 1 were replicated in Study 2. For example, the first contingency table indicates that there was a significant association between performance on the Destination Problem and the Drinking-age Problem. There was an excess of individuals in the correct/correct and incorrect/incorrect cells over what would be expected based on chance. Individuals in the correct/correct cell had higher SAT scores than the rest of the participants. Among participants responding incorrectly on the Destination Problem, there was no tendency for participants answering the Drinking-age Problem correctly to have higher SAT scores than those answering the Drinking-age Problem incorrectly (just as in Table 2). These basic trends were apparent in the remaining relationships

TABLE 7
Performance Relationships Among Several of the Problems in Study 2

	Drinking -age Problem Incorrect	*Drinking -age Problem Correct*
Destination Problem		
Incorrect	42 (38.6) [SAT = 1110]	228 (231.4) [SAT = 1095]
Correct	0 (3.4) [SAT = —]	24 (20.6) [SAT = 1190]

phi coefficient = .122 $\chi^2(1) = 4.36$, $P < .05$

	Cosmides Problem Incorrect	*Cosmides Problem Correct*
Destination Problem		
Incorrect	141 (134.1) [SAT = 1075]	129 (135.9) [SAT = 1122]
Correct	5 (11.9) [SAT = 1206]	19 (12.1) [SAT = 1186]

phi coefficient = .172 $\chi^2(1) = 8.69$, $P < .01$

	Margolis Problem Incorrect	*Margolis Problem Correct*
Destination Problem		
Incorrect	193 (184.6) [SAT = 1090]	77 (85.4) [SAT = 1117]
Correct	8 (16.4) [SAT = 1118]	16 (7.6) [SAT = 1227]

phi coefficient = .225 $\chi^2(1) = 14.83$, $P < .001$

	Drinking-age Problem Incorrect	*Drinking-age Problem Correct*
Abstract Problem		
Incorrect	39 (37.1) [SAT = 1112]	221 (222.9) [SAT = 1095]
Correct	3 (4.9) [SAT = 1077]	31 (29.1) [SAT = 1167]

phi coefficient = .056 $\chi^2(1) = 0.94$ *ns*

	Cosmides Problem Incorrect	*Cosmides Problem Correct*
Abstract Problem		
Incorrect	134 (129.1) [SAT = 1076]	126 (130.9) [SAT = 1122]
Correct	12 (16.9) [SAT = 1121]	22 (17.1) [SAT = 1180]

phi coefficient = .104 $\chi^2(1) = 3.17$, $P < .10$

	Margolis Problem Incorrect	*Margolis Problem Correct*
Abstract Problem		
Incorrect	184 (177.8) [SAT = 1088]	76 (82.2) [SAT = 1122]
Correct	17 (23.2) [SAT = 1122]	17 (10.8) [SAT = 1196]

phi coefficient = .143 $\chi^2(1) = 6.00$, $P < .025$

Figures in parentheses are the expected frequencies in that cell based on statistical independence, and figures in brackets are the mean SAT scores.

involving the Destination and Abstract Problems displayed in Table 7. The direction of the associations was the same in all cases. Specifically, participants responding with the P and not-Q choice on one problem were more likely to respond by choosing that pair on another problem. This was true even of associations between deontic and nondeontic problems where, as in Study 1, participants giving the modal response on the Drinking-age Problem (P and not-Q) were *more* likely to give the *non*-modal P and not-Q response on nondeontic problems.

STUDY 3

Despite much convergence between the results of Studies 1 and 2, the results of Study 1 might be considered suspect because only one order of administration was used (the five nondeontic problems always came first) and the order of problems in within-subject experiments has been shown to affect solution rates (Cox & Griggs, 1982; Platt & Griggs, 1993a). Furthermore, five consecutive nondeontic problems (with no intervening material) preceded the Drinking-age Problem in Study 1. In Study 3 we examine the cognitive ability correlates of performance on two tasks (a single nondeontic problem and the Drinking-age Problem from Study 1) in two groups of participants across which task order is counterbalanced. One group attempted the nondeontic problem first and the other group attempted the Drinking-age Problem first. An additional purpose of Study 3 was to test the generality of one key result from Study 2—the absence of a cognitive ability effect in a truly deontic version of the Drinking-age Problem. Thus, subsequent to the two counterbalanced problems in Study 3, each participant received a deontic selection task that is different from the Drinking-age Problem but which has shown robust facilitation in previous studies.

Method and Procedure

The participants were 215 undergraduate students recruited as in the previous two studies. The SAT scores of these participants will be reported as rescaled scores because the majority were derived from testings subsequent to the April 1995 rescaling of the SAT. Scores from testings prior to the April 1995 rescaling have been recentred according to ETS formulas. The mean reported SAT total score was 1194 (*SD* = 104), reasonably close to the institutional means of 1192 for 1994–95 and 1179 for 1995–96.

All participants completed both the Destination Problem and the Drinking-age Problem (without scenario) described in Study 1. In order A (completed by 109 participants) the Destination Problem immediately preceded the Drinking-age Problem, and in order B (completed by 106 participants) the Drinking-age Problem immediately preceded the Destination Problem.

Immediately after completing the Drinking-age and Destination Problems the participants in both groups completed the Sears Problem—a deontic selection task that has consistently displayed robust facilitation of performance (Dominowski, 1995; Griggs, 1983; Griggs & Cox, 1983). The version used in Study 3 was as follows:

> Suppose that you are the assistant manager at Sears, and it is your job to check sales receipts to make sure they are properly filled out according to a rule. The rule is: Any sale over $30 must be approved by the section manager, Mr. Jones. The amount of the sale is on one side of each receipt, and the space for the approval signature is on the other side. Which of the sales receipts shown below would you need to turn over in order to find out whether or not the rule is being violated?

$70	$22	Approval: Mr. Jones	Approval: . . . . . .
a = turn b = not turn	a = turn b = not turn	a = turn b = not turn	a = turn b = not turn

One participant failed to complete the Destination Problem and one participant failed to complete the Sears Problem, so *N* = 214 for these two tasks.

Results and Discussion

As in previous studies (see Cox & Griggs, 1982) task order did have a significant effect on solution rates in the Drinking-age Problem, suggesting that performance on this problem might have been depressed in Study 1 due to it following five consecutive nondeontic problems. Specifically, when the Drinking-age Problem was the first problem participants solved (order B), 76.3% answered it correctly; whereas when the Drinking-age Problem followed the Destination Problem (order A) only 58.1% of the participants answered it correctly, $\chi^2(1) = 8.66$, $P < .01$. However, even the order A solution rate was higher than that obtained in Study 1, suggesting that the additional nondeontic problems may have been depressing performance in Study 1. Order effects on the Destination Problem were much smaller and nonsignificant—6.8% answered correctly in order A and 9.7% in order B, $\chi^2(1) = 0.64$.

The SAT differences between correct and incorrect responders on both problems are presented in Table 8. The effect size (Cohen's *d)* displayed by the Destination Problem in Study 3 (.815) is between those displayed in Studies 1 and 2 (.502 and .859) and it was substantially larger than that displayed by the Drinking-age Problem (.347). More importantly, as in Study 1, the latter effect is disproportionately due to the performance of those individuals who *also* responded correctly to the Destination Problem. Among those responding incorrectly to the latter, there was only a nonsignificant 22-point advantage for correct Drinking-age responders.

The solution rate for the Sears Problem was quite high (59.3% responded correctly), consistent with previous findings of facilitation with this deontic problem (Dominowski, 1995; Griggs & Cox, 1983). The solution rate did not depend on the order of the previous two problems. Of the participants receiving

TABLE 8
Mean SAT Total Scores of Participants Who Gave the Correct and Incorrect Responses to the Three Selection Task Problems in Study 3

	Incorrect	*Correct*	*t value*	*Effect Size*[a]
Destination Problem	1187(197)	1270 (17)	3.21***	.815
Drinking-age Problem	1170 (72)	1206 (143)	2.39**	.347
Sears Problem	1189 (87)	1198 (127)	0.63	.088

Figures in parentheses are the number of participants giving the correct and incorrect responses.

The SAT scores in Study 3 are rescaled scores and are thus not directly comparable with the scores from Studies 1 and 2; df = 212 for the Destination and Sears Problems and 213 for the Drinking-age Problem

** = $P < .025$, *** = $P < .01$, all two-tailed

[a]Cohen's *d*

TABLE 9
Performance Relationships Among the Problems in Study 3

	Drinking -age Problem Incorrect	*Drinking -age Problem Correct*
Destination Problem		
Incorrect	71 (66.3) [SAT = 1173]	126 (130.7) [SAT = 1196]
Correct	1 (5.7) [SAT = 1020]	16 (11.3) [SAT = 1286]

phi coefficient = .173 $\chi^2(1) = 6.38$, $P < .025$

	Sears Problem Incorrect	*Sears Problem Correct*
Destination Problem		
Incorrect	85 (80.1) [SAT = 1187]	111 (115.9) [SAT = 1187]
Correct	2 (6.9) [SAT = 1275]	15 (10.1) [SAT = 1270]

phi coefficient = .174 $\chi^2(1) = 6.47$, $P < .025$

Figures in parentheses are the expected frequencies in that cell based on statistical independence, and figures in brackets are the mean SAT scores.

order A, 53.9% responded correctly on the Sears Problem and of the participants receiving order B, 62.0% responded correctly on the Sears Problem, a difference that was not statistically significant, $\chi^2(1) = 1.55$, ns. As Table 8 indicates, the difference in the mean SAT scores of those responding correctly on the Sears Problem and those responding incorrectly was not statistically significant.

The first contingency table displayed in Table 9 indicates that the relationships displayed in Studies 1 and 2 were replicated in Study 3. There was a significant association between performance on the Destination Problem and the Drinking-age Problem. There was an excess of individuals in the correct/correct and incorrect/incorrect cells and individuals in the correct/correct cell had higher SAT scores than the rest of the participants. The bivariate relationships between performance on the Destination and Sears Problems displayed the same patterns. In both cases, individuals responding with the P and not-Q choice on one problem were more likely to respond by choosing that pair on another problem.

Study 3 thus replicated the finding of Study 1 that the relation between responses and cognitive ability was much stronger for the nondeontic versions of the selection task than for the Drinking-age Problem without scenario. In both studies, participants answering the Drinking-age Problem correctly, but failing to answer a nondeontic version of the selection task correctly, failed to display significantly higher SAT scores. Study 3 established that this finding was not a function of preceding the Drinking-age Problem with five nondeontic problems as in Study 1. Additionally, Study 3 revealed that another truly deontic task (the Sears Problem) failed to display a difference in cognitive ability, thus replicating with a different problem the results of Study 2 from the Drinking-age Problem with scenario.

GENERAL DISCUSSION

As has been known for decades, people answering nondeontic problems correctly are few in number. Studies 1–3 give a fairly clear answer to the question of *who* answers nondeontic problems correctly. Individuals who solve such problems are disproportionately people of higher cognitive ability. These results (and those of Dominowski & Ansburg, 1996, discussed later) counter a longstanding assumption in the selection task literature that whether an individual answered correctly or not could not be predicted. For example, over 25 years ago Wason and Johnson-Laird (1972, p.173) declared that, regarding performance on the selection task, "only the rare individual takes us by surprise and gets it right. It is impossible to predict who he will be." Although agreeing with Wason and Johnson-Laird (1972) on just about nothing else, Wetherick (1971, p.213) agreed with them on the unpredictability of correct responding: "in Wason's experimental situation subjects do not choose the not-Q card nor do they stand and give three cheers for the Queen, neither fact is interesting in the absence of a plausible theory predicting that they should." Interestingly, Wetherick (1971, p.213) used the lack of data regarding individual differences on the task to skewer Wason's assumptions about what represents correct responding: "If it could be shown that subjects who choose not-Q are more intelligent or obtain better degrees than those who do not this would make the problem worth investigation, but I have seen no evidence that this is the case." The studies reported here in fact provide such evidence. The effect sizes of solving versus not solving a nondeontic task on SAT performance were in fact reasonably large, ranging from .502 to .859 and averaging .673 which is larger than moderate in most classifications of effect sizes (Rosenthal & Rosnow, 1991, p.446).

In contrast, the pattern of associations with cognitive ability on the Drinking-age and Sears Problems is entirely different from that on nondeontic problems. The version of the Drinking-age Problem that used the full policeman scenario (Study 2) resulted in very high solution rates, as in previous studies (Griggs & Cox, 1982; Pollard & Evans, 1987), but the individuals who solved the problem were no higher in cognitive ability than those who did not. Likewise, the individuals who solved the Sears Problem in Study 3 were no higher in cognitive ability than those who did not.

A more difficult version of the Drinking-age Problem, without the policeman scenario (used in Studies 1 and 3), did produce a difference in SAT scores that attained significance—but when participants who had also solved a nondeontic problem were eliminated, the difference in SAT scores disappeared. That is, when only those participants answering nondeontic problems incorrectly were considered, there was no difference in SAT scores between solvers and nonsolvers of the Drinking-age Problem. This finding is consistent with the conjecture that there might be a small number of individuals who respond nonheuristically to the nondeontic problems (see Evans, 1984, 1989, 1996; Evans & Clibbens, 1995) and it might be assumed that these individuals might also respond analytically on the Drinking-age Problem. But when these people are eliminated, it cannot be assumed that the remaining solvers of the Drinking-age Problem are responding analytically at all. The bivariate relationships that were examined also supported this conjecture. When performance on the Abstract or Destination Problems was examined in conjunction with performance on the Drinking-age, Sears, or Cosmides Problems, there was consistently an over-representation of individuals in the correct/correct cell.

Across both studies and across all problems, there was a consistent tendency for the incorrect response P to be associated with high SAT scores. This finding might be interpreted as indicating that Margolis (1987) is correct in stating that some individuals develop an open reading of the task (in terms of choosing categories) rather than a closed reading (in terms of choosing cards) and that this is a reasonable construal (see also Lowe, 1993; Nickerson, 1996) given that it is one that is attractive to the more cognitive able individuals. Margolis (1987; see Nickerson, 1996) has shown that if the participant views the task from an open scenario then P-only is a reasonable choice. It is thus intriguing that the individuals opting for the P choice had, with the exception of the correct responders, the highest SAT scores. It is also interesting that the difference between P responders and matching (P and Q) responders was fairly large on the Abstract (64 points), Destination (55 points), Cosmides (75 points), and Drinking-age (112 points) Problems, but was smallest on the Margolis Problem (13 points) which was explicitly designed to make the open reading less plausible. The mean SAT score of the P responders on the Margolis Problem (1109) was substantially below that of P responders on the Abstract Problem (1153) where the open reading is perhaps most plausible.

The Margolis Problem itself displayed an effect size for cognitive ability (.410) that was intermediate in size between that of the Drinking-age Problems and Sears Problem (.229, –.050, .347, and .088) and all of the nondeontic problems (which displayed effect sizes largely between .550 and .850). We believe that the Margolis task is a hybrid problem—interpreted as indicative by most subjects, but interpreted as deontic by a substantial minority. Its hybrid status derives from the fact that it has no features emphasising cost/benefit social exchange trade-offs, nor does it emphasise the concept of cheater detection.

However, it does contain the term “must” and contains violation instructions that in conjunction with ”must” might help to trigger a deontic interpretation in some participants. Thus, under the assumption that it is solution under an indicative construal that is the cause of cognitive ability differences and that a deontic construal serves to dilute cognitive ability differences (see later discussion), the intermediate-sized cognitive ability effect is consistent with the hybrid nature of the problem.

Also requiring explanation is the outcome on the Cosmides Problem where we also observed a cognitive ability effect of intermediate size (.469). Again, it may be that we used a version of this problem that was ambiguous in its status—evoking an indicative interpretation in some subjects and a deontic interpretation in others. For example, Platt and Griggs (1993a, pp.167–168)—following Manktelow and Over (1991)—argued that “subjects’ underlying understanding of deontic terms helps them to reason well when the content of the rule is of real importance to them.” They found that the presence of cost–benefit information (missing from our version of the problem) was very important in triggering a deontic social contract interpretation. Two cues present in our version (the cheating perspective and the presence of “must”) were less consistently associated with the deontic interpretation in their study. Thus, as with the Margolis Problem, the Cosmides Problem we used may have elicited both deontic and indicative interpretations. If it is only the latter that is associated with cognitive ability differences (see later discussion) then the hybrid nature of this problem is consistent with an intermediate effect size for cognitive ability.

It should be noted that the magnitude of the effect of cognitive ability across the various problems is not just tracking the overall difficulty of the task. Although it is true that the magnitude of the effect in the hardest tasks (Abstract, Destination) is vastly larger than that for the easiest (Drinking-age Problem with scenario), across the three studies there were numerous strong reversals of this relationship. For example, the Drinking-age Problem in Study 1 was equal in difficulty to the Margolis Problem (31.5% correct versus 31.6% correct) and actually more difficult than the Cosmides Problem (50.3% correct) but it displayed an SAT difference smaller than either of these two problems (effect size of .229 versus .410 and .469, respectively). Likewise, the Sears Problem was virtually as difficult as the Cosmides Problem (59.3% correct versus 50.3% correct), but the Sears Problem displayed a substantially lower SAT difference (effect size of .088 versus .469). The solution rate for the Drinking-age Problem in Study 1 (31.5%) was substantially lower than that for the Drinking-age Problem in order B of Study 3 (76.3%) but the SAT effect size was actually lower in the former than the latter (.229 vs .485). An analysis of the two orders of the Margolis Problem further reinforces the conclusion that the effect of cognitive ability across the various problems is not solely a function of task difficulty.

Just as difficulty does not explain the cognitive ability differences, transfer effects cannot explain all of the associations between deontic and nondeontic tasks. For example, neither the Destination Problem nor the Sears Problem in Study 3 displayed an order effect, yet performance on these two problems was significantly associated. Additionally, in the condition of Study 3 where the Drinking-age Problem was presented first, and the Destination Problem (which displayed no order effect) was presented second, there was still a significant association. Several examples such as this occur throughout the studies and suggest that the excess of subjects in the correct/correct cell is not solely due to transfer effects.

Finally, previous studies of individual differences on selection task performance have examined the effects of educational level (Hoch & Tschirgi, 1985; Jackson & Griggs, 1988), area of expertise (Jackson & Griggs, 1988), and scientific background (Griggs & Ransdell, 1986; Tweney & Yachanin, 1985). Most of these investigations employed fairly small samples and none of them contained a systematic investigation of the effects of cognitive ability. It is thus difficult to relate this work to the present findings. For example, the finding of Hoch and Tschirgi (1985) that abstract selection task performance was related to educational level, and that of Tweney and Yachanin (1985) that the performance of scientists was superior, might be viewed as consistent with our observed relationships with cognitive ability. However, both of the former findings have failed to replicate in some small-sample studies (Griggs & Ransdell, 1986; Jackson & Griggs, 1988).

CONCLUSION

Our results on individual differences in cognitive ability associated with differing responses to selection task problems can be parsimoniously accommodated within the dual process framework of Evans (1984, 1989, 1996; Evans & Over, 1996) and the optimal data selection model of Oaksford and Chater (1994). The latter model correctly predicts that the modal bivariate performance relationship on a clearly deontic and clearly nondeontic task should be correct (P and not-Q) and incorrect (P and Q), respectively (see Table 9). However, a slight modification of the dual process view (Evans, 1996; Evans & Over, 1996) might also explain the tendency (also displayed in Table 9) for the correct/correct cell to be overpopulated. We would argue that the significant deviation from statistical independence that consistently tends in this direction is due to the existence of a subset of individuals who *are* viewing the nondeontic task as a problem of deductive logic and are reasoning analytically to their choices.

Part of the title of a recent article (Evans, 1996)—"deciding before you think"—sums up this framework. Evans (1996, p.224) argues that abstract or other nondeontic versions of the selection task "may not be a reasoning task at all, in the sense that it may fail to elicit any cognitive process of the type we would wish to describe as reasoning." If we view the responses of the vast majority of incorrect responders in nondeontic versions as dominated by a preconscious relevance judgement (see Evans & Over, 1996, for the most recent discussion of the dual-process view), we are still left with the small proportion of correct responders. Are these individuals random variation (as the earlier quotes from Wetherick and Wason & Johnson-Laird seem to assume), or might these be individuals who are approaching the task in a more analytic manner? Perhaps instead of an analytic stage dominated by attempts to justify the early heuristic attentional responses, some individuals use the analytic processing stage to critically examine the implications of turning *all* of the cards—not just the ones brought to attention through a preconscious relevance judgement. That is, some individuals might actually think before they decide. And if we view intelligence as encompassing important self-regulatory and metacognitive components (Byrnes, 1995; Klaczynski & Gordon, 1996; Sternberg, 1985) then we might well expect the individuals who think before they decide to be individuals of higher cognitive ability.

In short, what the results may be telling us is that in the abstract task, and in various versions of nondeontic problems with thematic content, a considerable amount of thinking is required before deciding—because the early heuristic relevance judgements must be overcome. However, in the Drinking-age Problem, it is possible to be correct by deciding without thinking because the early relevance judgements cue the correct response. With this rule, it is fine to think only after you decide because preconscious heuristics lead to the correct response.

The heuristic-analytic framework fits particularly well with the present results because Evans and Over (1996) have linked their notion of preconscious, heuristic processing with work in the implicit learning tradition (Reber, 1993). Specifically, they endorse Reber's (1993) view of such preconscious processes as evolutionarily more ancient, more robust in the face of insult, less variable, and less related to intelligence than conscious processes. The latter assumption, for which there is some empirical evidence (McGeorge, Crawford, & Kelly, 1997; Reber, Walkenfeld, & Hernstadt, 1991), is the crucial one for the purposes of explaining the cognitive ability differences observed in the present study. In nondeontic selection tasks, the preconscious relevance response cues a response (P and Q) that is different from the one that will be deemed correct under a deductive construal. The latter will result from conscious, analytic processing overcoming the heuristic cueing and because the conscious mechanism is related to intelligence, individuals adopting this construal will tend to be of higher cognitive ability.

In contrast, in deontic problems, both deontic and rule-based logics are cueing construals of the problem that dictate the same response (P and not-Q). Whatever one's theory of responding in deontic tasks—preconscious relevance judgements, pragmatic schemas, Darwinian algorithms, or evolutionarily maximised expected utility—the mechanisms are likely tacit in the sense of Evans and Over (1996) and Reber (1993) and thus are unlikely to be strongly

TABLE 10
Simulation of Differences in Mean SAT Scores Assuming the Heuristic-analytic Account of Individual Differences

Actual Data from Abstract Task of Study 2

	Incorrect		*Correct*	
Heuristic 1 (Linguistic Relevance; see Evans, 1995) Response = P&Q	1091	(140)		
Analytic Failure (P; All; P, Q; Other)	1107	(120)		
Analytic Success P and not-Q			1159	(34)
Actual Mean	1098		1159	

Simulated Drinking-age Problem (With Scenario)
Results Based on a Total Heuristic Change

	Incorrect		*Correct*	
Heuristic 1 (Linguistic Relevance; see Evans, 1995) Response = P&Q	–			
Heuristic 2 (Pragmatic Relevance; see Evans, 1995) Response = P & not-Q			1091	(140)
Analytic Failure (P; All; P, Q, not-Q; Other)	1107	(120)		
Analytic Success P and not-Q			1159	(34)
Simulated Mean	1107		1104	
Actual Mean	1110		1104	

Actual data from nondeontic tasks of Study 1:

	Incorrect		*Correct*	
Heuristic 1 (Linguistic Relevance; see Evans, 1995) Response = P&Q	1104	(150)		
Analytic Failure (P; All; P, Q, not-Q; Other)	1094	(161)		
Analytic Success P and not-Q			1185	(22)
Actual Mean	1099		1185	

Simulated Data for the Drinking-age Problem of Study 1 (Without Scenario):

	Incorrect		*Correct*	
Heuristic 1 (Linguistic Relevance; see Evans, 1995) Response = P&Q	1104	(75)		
Heuristic 2 (Pragmatic Relevance; see Evans, 1995) Response = P& not-Q			1104	(75)
Analytic Failure (P; All; P, Q, not-Q; Other)	1094	(161)		
Analytic Success P and not-Q			1185	(22)
Simulated Mean	1097		1122	
Actual Mean	1099		1124	

Number of participants in parentheses

associated with analytic intelligence. Hence, such processes will draw subjects of *both* high and low analytic intelligence to the same response dictated by the rule-based system and thus will serve to dilute cognitive ability differences between correct and incorrect responders.

An informal simulation of our conceptualisation is presented in Table 10. Following Evans (1995), we propose that the matching response (P and Q) is triggered by a preconscious linguistic heuristic. The top of Table 10 indicates that, on the Abstract Problem of Study 2, the 140 individuals who gave this response had mean SAT scores of 1091. The 34 individuals choosing P and not-Q had a mean SAT score of 1159. These are assumed to be individuals who reasoned analytically to the correct solution. As a simplifying assumption, the individuals in the remaining response categories (P; All; P, Q, not-Q; Other) are assumed to be individuals for whom analytic processing also overrode heuristic processing but who did not analytically reason to the correct solution. These individuals had mean SAT scores higher than the heuristic responders (1107 versus 1091), but lower than the correct responders. The sizeable number of analytic processors who did not solve the problem is consistent with O'Brien's (1995) argument that for a mental logic without direct access to the truth table for the material conditional, the selection task is a very hard problem. This subgroup, when combined with matching responders, yields an overall SAT mean for the incorrect responders of 1098—which is 61 points below the mean of the correct responders.

The next analysis in the Table simulates how the difference between correct and incorrect responders would change upon presentation of a clearly deontic problem such as the Drinking-age Problem with scenario. Two strong assumptions are made, but moderate relaxation of either assumption leaves the essential lesson to be drawn from the simulation unchanged. The first assumption is that the heuristic processors remain heuristic processors, but the relevance cue is now pragmatic rather than linguistic (see Evans, 1995) and in the context of the Drinking-age Problem cues the correct response (P & not-Q). The second assumption is that the analytic processors remain analytic processors. The simulated means indicate that the shift in heuristics has totally eliminated the difference in mean SAT scores between the groups of correct and incorrect responders. It is also easy to see that a relaxation of the second assumption (that the analytic processors remain analytic processors) would have virtually no effect on this pattern.

The next two analyses in Table 10 illustrate the expected outcome when an only partially effective deontic manipulation is attempted—the Drinking-age Problem without scenario of Study 1. The first indicates that, on the five nondeontic problems from Study 1 (see Table 3), the 150 individuals who gave the matching response had mean SAT scores of 1104. The 22 individuals choosing P and not-Q had mean SAT scores of 1185. The individuals in the remaining response categories (P; All; P, Q, not-Q; Other) had mean SAT scores of 1094. This subgroup, when combined with matching responders, yields an overall SAT mean for the incorrect responders of 1099—which is 86 points below the mean of the correct responders.

The next analysis in the Table simulates how the difference between correct and incorrect responders would change upon presentation of the Drinking-age Problem without scenario used in Study 1. As previously, it is assumed that analytic processors remain analytic processors. It is also assumed that, without scenario, this Drinking-age Problem is successful in converting only one half of the heuristic processors from linguistically based relevance cues to pragmatically based relevance cues. The simulated means in the Table indicate that the shift in heuristics results in a drop in the difference between means from 86 points to 25 points—precisely the difference observed in Study 1 (and slightly less than the 36-point difference observed in Study 3).

In short, we posit that correct responders on nondeontic tasks are largely analytic processors; whereas the group of correct responders on deontic tasks such as the Drinking-age and Sears Problems comprises a small number of analytic processors and a much larger group of heuristic processors. Once the former are removed from the group of correct responders, the latter are no more likely to be of high cognitive ability than are those who gave an incorrect response. Consistent with Evans and Over's (1996) and Reber's (1993) speculations about individual differences, it is responses determined by analytic processing that create differences in cognitive ability. In contrast, correct responses determined by heuristic processes serve to dilute such differences. [Note that simulating the outcome for the Margolis Problem would be much more complicated because it is a hybrid problem. Some of the facilitation might come from changing heuristics; some might come from cues that make more of the subjects process analytically rather than heuristically; and some facilitation might come from cues that make it easier for those who do reason analytically to compute the appropriate response.]

Champions of human rationality have been at pains to defend the modal response of P and Q on abstract tasks and P and not-Q on deontic tasks as arising from the same basic and rational thought processes (Wetherick, 1995). The pragmatic reasoning theories and social exchange theories appear problematic for this view. As Dominowski and Ansburg (1996, p.5) argue:

> although the formally-correct cards are selected, there might be no reasoning taking place—i.e., the person might simply have learned what to do in the situation presented. At best, success might reflect narrow reasoning processes strictly tied to a particular kind of content. No content-free reasoning processes are assumed to play a role. In short, these accounts characterize successful performance on thematic problems as rather isolated behavior.

This is precisely Wetherick's (1995, p.439) concern. He rejects the

> theoretical superstructures that have been erected to explain apparently 'correct' performance on these aberrant versions of the four-card task [because the

> pragmatic and social exchange theories are] irrelevant. Ordinary logic is sufficient to show that, in every case, intelligent subjects do what intelligent subjects would be rationally expected to do.

Wetherick (1995) seeks support in some unpublished findings reported by Dominowski and Dallob (1991; see also Dominowski & Ansburg, 1996). They reported that performance on abstract and thematic tasks was correlated and that performance on both was related to several general reasoning tests. Wetherick (1995) argues that these results contradict the view that different mental mechanisms are involved in the abstract and deontic versions of the task. Instead, Wetherick (1995, p.440) views the results as supporting the notion that both types of task "call on the same mental processes. I would argue that the preceding argument of this section suggests strongly that these processes are the processes of ordinary logic."

But Wetherick's (1995) citing of Dominowski and Dallob's study in this context is puzzling because that study (as did ours) revealed correlations with what is *typically* considered to be the normative response on the abstract task (P and not-Q)—not with the response Wetherick considers to be normative and rational (P and Q; in fact, he has argued that people who fail to give the P and not-Q response "are to be congratulated," Wetherick, 1970, p.214). Similarly, in both studies, it was the P and not-Q responders who displayed higher cognitive ability, not the P and Q responders.

There were some differences between Dominowski and Ansburg's (1996) results and ours, however. We became aware of their unpublished work just prior to writing up the present studies. As in our studies, they found correlations between performance on thematic and nonthematic problems. Like us, Dominowski and Ansburg (1996) found that measures of cognitive ability were correlated with performance on problems with familiar content but arbitrary rules. Unlike our studies, they found cognitive ability differences on their thematic problems that were virtually as strong as those on their nonthematic problems. However, there are numerous differences between their materials and ours that might explain this difference. First, they did not investigate the Drinking-age Problem, which has been one of the most consistent deontic problems. Additionally, no violation instructions were included in the thematic problems and the deontic "must" was used in all of the nonthematic problems. One of the thematic problems was the Stamp Problem (if a letter is sealed, then it must carry a 20-cent stamp) which has been a very inconsistent producer of facilitation (Manktelow & Evans, 1979). These factors all served to decrease the differences between the arbitrary and thematic problems. That the differences were minimised is clearly indicated by the fact that the solution rates on the arbitrary and thematic problems were quite similar (36% and 47%, respectively). Thus, the processing mechanisms engaged by their two types of problems are likely to be much more similar than those engaged by the problems studied here.

The results of our studies suggest a refinement in the theories of optimal data selection that assume that people view the task as an inductive problem of probabilistic hypothesis testing (e.g. Evans & Over, 1996; Nickerson, 1996; Oaksford & Chater, 1994). Such theories may well account for the performance of the majority of individuals. However, our results suggest that perhaps 10% of the participants—disproportionately those of higher cognitive ability—do view the task as a deductive problem, do reason analytically, and thus will display performance patterns that will not be well fitted by these theories. We have shown that who these individuals are can be predicted by their performance on other nondeontic problems and by their general cognitive ability. For example, in Study 1, participants who were above the median in cognitive ability and who answered any two nondeontic problems correctly had an over 85% chance of answering a further nondeontic problem correctly. Models of optimal data selection should fit selection task results better if these individuals are eliminated from the analysis, because individuals of higher cognitive ability may be more likely to override evolutionary optimised computations in order to pursue a normative solution.

Manuscript received 28 July 1997
Revised manuscript received 11 February 1998

REFERENCES

Ackerman, P.L., & Heggestad, E.D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*, 219–245.

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

Anderson, J.R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*, 471–517.

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

Byrnes, J.P. (1995). Domain-specificity and the logic of using general ability as an independent variable or covariate. *Merrill-Palmer Quarterly*, *41*, 1–24.

Cacioppo, J.T., Petty, R.E., Feinstein, J., & Jarvis, W. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*, 197–253.

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Caryl, P.G. (1994). Early event-related potentials correlate with inspection time and intelligence. *Intelligence* , *18*, 15–46.

Chater, N., & Oaksford, M. (1996). Deontic reasoning, modules and innateness: A second look. *Mind & Language*, *11*, 191–202.

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

Cheng, P.W., Holyoak, K.J., Nisbett, R.E., & Oliver, L.M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology*, *18*, 293–328.

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

Cox, J.R., & Griggs, R.A. (1982). The effects of experience on performance in Wason's selection task. *Memory & Cognition*, *10*, 496–502.

Cummins, D.D. (1996). Evidence for the innateness of deontic reasoning. *Mind & Language*, *11*, 160–190.

Deary, I.J., & Stough, C. (1996). Intelligence and inspection time. *American Psychologist*, *51*, 599–608.

Detterman, D. (1994). Intelligence and the brain. In P.A. Vernon (Ed.), *The neuropsychology of individual differences* (pp.35–57). San Diego, CA: Academic Press.

Dominowski, R.L. (1995). Content effects in Wason's selection task. In S.E. Newstead & J. St. B. T. Evans (Eds.), *Perspectives on thinking and reasoning* (pp.41–65). Hove, UK: Erlbaum.

Dominowski, R.L., & Ansburg, P. (1996). *Reasoning abilities, individual differences, and Wason's selection task.* Manuscript under review.

Dominowski, R.L., & Dallob, P. (1991, September). *Reasoning abilities, individual differences, and the four card problem*. Paper presented at the Eighth Annual Conference, Cognitive Psychology Section, British Psychological Association, Oxford.

Evans, J.St.B.T. (1972). Interpretation and matching bias in a reasoning task. *Quarterly Journal of Experimental Psychlogy*, *24*, 193–199.

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

Evans, J.St.B.T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, *75*, 451–468.

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

Evans, J.St.B.T. (1995). Relevance and reasoning. In S.E. Newstead & J.St.B.T. Evans (Eds.), *Perspectives on thinking and reasoning* (pp.147–171). Hove, UK: Erlbaum.

Evans, J.St.B.T. (1996). Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology*, *87*, 223–240.

Evans, J.St.B.T., & Clibbens, J. (1995). Perspective shifts on the selection task: Reasoning or relevance? *Thinking and Reasoning*, *1*, 315–323.

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., Newstead, S.E., & Byrne, R.M.J. (1993). *Human reasoning: The psychology of deduction*. Hove, UK: Erlbaum.

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

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

Griggs, R.A. (1983). The role of problem content in the selection task and in the THOG problem. In J.*S*.B.T. Evans (Ed.), *Thinking and reasoning: Psychological approaches* (pp.16–43). London: Routledge & Kegan Paul.

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

Griggs, R.A. (1989). To " see" or not to "see": That is the selection task. *Quarterly Journal of Experimental Psychology*, *41*A, 517–529.

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

Griggs, R.A., & Cox, J.R. (1983). The effects of problem content and negation on Wason's selection task. *Quarterly Journal of Experimental Psychology*, *35*, 519–533.

Griggs, R.A., & Cox, J.R. (1993). Permission schemas and the selection task. *Quarterly Journal of Experimental Psychology*, *46*A, 637–651.

Griggs, R.A., & Ransdell, S.E. (1986). Scientists and the selection task. *Social Studies of Science*, *16*, 319–330.

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

Holyoak, K.J., & Cheng, P.W. (1995a). Pragmatic reasoning with a point of view. *Thinking and Reasoning*, *1*, 289–313.

Holyoak, K.J., & Cheng, P.W. (1995b). Pragmatic reasoning from multiple points of view: A response. *Thinking and Reasoning*, *1*, 373–389.

Horn, J.L., & Cattell, R.B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica* , *26*, 1–23.

Hunt, E. (1987). The next word on verbal ability. In P.A. Vernon (Ed.), *Speed of information-processing and intelligence* (pp.347–392). Norwood, NJ: Ablex.

Jackson, S.I., & Griggs, R.A. (1988). Education and the selection task. *Bulletin of the Psychonomic Society*, *26*, 327–330.

Jackson, S.I., & Griggs, R.A. (1990). The elusive pragmatic reasoning schemas effect. *Quarterly Journal of Experimental Psychology*, *42*A, 353–373.

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

Klaczynski, P., & Laipple, J. (1993). Role of content domain, logic training, and IQ in rule acquisition and transfer. *Journal of Experimental Psychology: Learning*, *Memory and Cognition*, *19*, 653–672.

Klaczynski, P.A., & Gordon, D.H. (1996). Self-serving influences on adolescents' evaluations of belief-relevant evidence. *Journal of Experimental Child Psychology*, *62*, 317–339.

Liberman, N., & Klar, Y. (1996). Hypothesis testing in Wason's selection task: Social exchange cheating detection or task understanding. *Cognition*, *58*, 127–156.

Lowe, E.J. (1993). Rationality, deduction and mental models. In K. Manktelow & D. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp.211–230). London: Routledge.

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.

Manktelow, K.I., & Over, D.E. (1990). Deontic thought and the selection task. In K. Gilhooly, R. Keane, R. Logie, & G. Erdos (Eds.), *Lines of thinking (Vol. 1)* (pp.153–164). Chichester, UK: John Wiley.

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

Manktelow, K.I., Sutherland, E., & Over, D.E. (1995). Probabilistic factors in deontic reasoning. *Thinking and Reasoning*, *1*, 201–219.

Margolis, H. (1987). *Patterns*, *thinking, and cognition*. Chicago: University of Chicago Press.

Matarazzo, J.D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th Edn.). Baltimore: The Williams & Wilkins Co.

McGeorge, P., Crawford, J., & Kelly, S. (1997). The relationships between psychometric intelligence and learning in an explicit and an implicit task. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *23*, 239–245.

Newstead, S.E., & Evans, J.St.B.T. (Eds.). (1995). *Perspectives on thinking and reasoning*. Hove, UK: Erlbaum.

Nickerson, R.S. (1996). Hempel's paradox and Wason's selection task: Logical and psychological puzzles of confirmation. *Thinking and Reasoning*, *2*, 1–31.

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. (1995). Theories of reasoning and the computational explanation of everyday inference. *Thinking and Reasoning*, *1,* 121–152.

O'Brien, D.P. (1995). Finding logic in human reasoning requires looking in the right places. In S.E. Newstead & J.*S*.B.T. Evans (Eds.), *Perspectives on thinking and reasoning* (pp.189–216). Hove, UK: Erlbaum.

Over, D.E., & Manktelow, K.I. (1995). Perspectives, preferences, and probabilities. *Thinking and Reasoning, 1*, 364–371.

Platt, R.D., & Griggs, R.A. (1993a). Darwinian algorithms and the Wason selection task: A factorial analysis of social contract selection task problems. *Cognition*, *48*, 163–192.

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

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

Pollard, P., & Evans, J.St.B.T. (1987). Content and context effects in reasoning. *American Journal* of *Psychology, 100*, 41–60.

Reber, A.S. (1993). *Implicit learning and tacit knowledge*. New York: Oxford University Press.

Reber, A.S., Walkenfeld, F.F., & Hernstadt, R. (1991). Implicit and explicit learning: individual differences and IQ. *Journal of Experimental Psychology: Learning*, *Memory and Cognition*, *17*, 888–896.

Rosenthal, R., & Rosnow, R.L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd Edn). New York: McGraw-Hill.

Sperber, D., Cara, E, & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, *57*, 31–95.

Sternberg, R.J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* Cambridge: Cambridge University Press.

Straughn, C.S., & Straughn, B.L. (Eds.) (1995). *Lovejoy's college guide* (23rd Edn.). New York: Simon & Schuster.

Tweney, R.D., & Yachanin, S. (1985). Can scientists rationally assess conditional inferences? *Social Studies of Science*, *15*, 155–173.

Vernon, P.A. (1993). *Biological approaches to the experimental of human intelligence.* Norwood, NJ: Ablex.

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

Wason, P.C. (1969). Regression in reasoning? *British Journal of Psychology*, *60*, 471–480.

Wason, P.C., & Evans, J.St.B.T. (1975). Dual processes in reasoning? *Cognition*, *3*, 141–154.

Wason, P.C., & Johnson-Laird, P.N. (1972). *Psychology of reasoning*. Cambridge, MA: Harvard University Press.

Wetherick, N.E. (1970). On the representativeness of some experiments in cognition. *Bulletin of the British Psychological Society*, *23*, 213–214.

Wetherick, N.E. (1971). Representativeness in a reasoning problem: A reply to Shapiro. *Bulletin of the British Psychological Society*, *24*, 213–214.

Wetherick, N.E. (1995). Reasoning and rationality: A critique of some experimental paradigms. *Theory & Psychology*, *5*, 429–448.

Willingham, W.W., Lewis, C., Morgan, R., & Ramist, L. (1990). *Predicting college grades: An analysis of institutional trends over two decades*. Princeton, NJ: Educational Testing Service.

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