Can tutoring improve performance on a reasoning task under deadline conditions?

MAGDA OSMAN
University College London, London, England

The present study examined the effectiveness of a tutoring technique that has been used to identify and address participants' misunderstandings in Wason's selection task. In particular, the study investigated whether the technique would lead to improvements in performance when the task was presented in a deadline format (a condition in which time restrictions are imposed). In Experiment 1, the effects of tutoring on performance were compared in free time (conditions in which no time restrictions are imposed) and deadline task formats. In Experiment 2, improvements in performance were studied in deadline task formats, in which the tutoring and test phases were separated by an interval of 1 day. The results suggested that tutoring improved performance on the selection task under deadline and in free time conditions. Additionally, the study showed that participants made errors because they had misinterpreted the task. With tutoring, they were able to modify their initial misunderstandings.

Wason's (1966) selection task is one of the most intensively studied problems used to examine the psychological processes involved in reasoning. In the standard version of the task, people are required to assess the truth of a conditional statement ("if *p*, then *q*"). The most commonly known example is the vowel–even number problem, in which the conditional statement takes this form: If a card has a vowel on one side (antecedent), then it has an even number on the other side (consequent). Along with this statement, participants receive information about the visible sides of four cards, two of which are letter cards (e.g., E and K) and two number cards (e.g., 4 and 7). Participants are then instructed to select, on the basis of the information presented on the visible side, which card or cards they need to turn over to test the statement.

Typically, in arbitrary versions such as the vowel–even number problem, only a few participants—4% in Johnson-Laird and Wason (1970)—select the logically correct values P and ¬Q—that is, the vowel and the odd number cards. However, there has been considerable debate whether logic is the appropriate normative standard by which to assess participants' performance (for a contrasting view, see Oaksford & Chater, 1994). The most common choice (e.g., 46% in Johnson-Laird & Wason, 1970) is to choose the P and Q values, the vowel and the even number cards; the next most common choice (33% in Johnson-Laird & Wason, 1970) is to choose the P value, the vowel card.

The poor performance found on standard abstract tasks contrasts with the dramatic improvements reported in studies (e.g., 74% in Griggs & Cox, 1982) that include deontic rules, which are rules about permission, such as social agreements or moral laws (e.g., Cheng & Holyoak, 1985; Cosmides, 1989; Gigerenzer & Hug, 1992; Griggs & Cox, 1982). However, there is evidence to suggest that nondeontic versions also facilitate performance (Almor & Sloman, 1996; Gebauer & Laming, 1997), and that deontic versions are simply other examples of a type of format in which facilitation occurs.

A variety of tutoring techniques has been developed to improve performance on arbitrary versions of the selection task. The techniques used in early tutoring studies were designed to increase selections of the ¬Q card and reduce selections of the Q card by broadening participants' representations of the task through logical insight (Johnson-Laird & Wason, 1970; Wason, 1968, 1969; Wason & Golding, 1974; Wason & Johnson-Laird, 1970; Wason & Shapiro, 1971). The tutoring techniques included teaching people the logical structure of conditionals; asking participants to imagine turning over the cards that would render the statement false; and asking people to consider separately only the cards that would prove the statement true. While the techniques did not lead to an increase in the correct card choices, Wason (1969; Wason & Shapiro, 1971) found that these techniques attenuated the frequency with which the most commonly selected erroneous card was chosen.

Later tutoring studies focused on facilitating correct performance by modifying the context in which the task is framed. Cheng, Holyoak, Nisbett, and Oliver (1986) investigated whether teaching logic is sufficient to improve performance on the selection task. In their study, they compared three different tutoring procedures: (1) a seven-page logic tutorial on conditional reasoning; (2) a tutorial

M. Osman, m.osman@ucl.ac.uk

on deontic versions of the selection task with feedback on performance; and (3) a combination of (1) and (2). After tutoring, participants were presented with arbitrary and deontic versions of the selection task. Cheng et al. found that only the combined condition improved performance in all versions of the task. In the other treatments, improvement in performance depended on the context. The logic tutorial improved performance only on arbitrary problems, and the deontic examples improved performance only on deontic problems. Cheng et al. also compared performance on the task before and after students had taken part in a 12-week introductory course on logic. They found that this course did not lead to improved performance in arbitrary and deontic versions of the selection task.

Recently, tutoring techniques have been tailored to typical misunderstandings of the statement (e.g., Margolis, 2000; Stenning & van Lambalgen, 2001). Current evidence suggests that, in arbitrary versions, the range of card combinations selected reflects different understandings of the conditional statement (e.g., Gebauer & Laming, 1997; Osman & Laming, 2001; Stenning & van Lambalgen, 1999). The most common misunderstandings are (1) to misread *one side/other side* as *top/bottom*, which leads to the selection of the P card because, in the vowel–even number example, participants assume that by turning over the A card they will see an even number on its underside; (2) the conditional as a biconditional (i.e., "if P then Q, if Q then P"), leading to the selection of all four cards; and (3) a combination of (1) and (2), leading to the selection of the cards P and Q (Gebauer & Laming, 1997; Osman & Laming, 2001).

Stenning and van Lambalgen (1999, 2004) revealed the numerous misunderstandings that affect how people evaluate the unseen sides of the example cards, and what bearing they think this information has on the validity of the statement. Stenning and van Lambalgen (2001, 2004) used a Socratic tutoring technique in which participants were asked probing questions to help them examine their own understanding of the task. Participants whose misunderstandings were thereby revealed were tutored to assist them in discovering the correct solution for themselves. Performance increased in standard arbitrary problems and transferred to a complex conditional task that involved falsifying two statements simultaneously (e.g., "If vowel, then even number"; "If consonant, then even number"). Stenning and van Lambalgen (2001, 2004) concluded that people possess deductive competence but are unable to utilize this ability because they reason from an inaccurate representation of the conditional statement and the cards. Tutoring is a way of constraining the possible range of misinterpretations, and it guides people to a logical interpretation from which they can reason correctly.

Why are some tutoring techniques more successful than others? The studies that show successful tutoring techniques reduce the sources of difficulty in interpreting standard abstract versions of the task. Tutoring in standard logic alone is not sufficient to induce sustained insight into the task, but it is a useful method of detecting people's misunderstandings of the statement. People need to anchor the knowledge that they acquire during tutor-

ing, in order to understand how and when to use it (Cheng et al., 1986). By reorienting people's existing knowledge (Stenning & van Lambalgen, 2001, 2004), successful tutoring procedures help achieve to this. Socratic-style tutoring methods are informative because, through judicious questioning, they help raise people's awareness of the problems they have with the task, and encourage them to discover solutions.

Dual system reasoning theorists (Evans & Over, 1996; Sloman, 1996; Stanovich & West, 2000) have argued that the limited success of logic-based tutoring, especially compared with context-dependent tutoring, occurs because reasoning comprises two functionally distinct systems: System 1 (heuristic processing) and System 2 (analytic processing). The processes characteristic of these systems have distinct roles, differ according to the type of information encoded, vary according to the level of expressible knowledge, and result in different responses (see Osman, 2004). The central proposition of the dual system framework of reasoning is that reasoning begins with System 1, which is inaccessible to conscious awareness and involves heuristic processing (the function of heuristics is to bias attention towards particular properties of the problem space). System 2 occurs in the later stages of reasoning and involves analytical processes that involve a type of deliberate and explicit thinking, such as making inferences from possibilities that may be abstract or have no referent in the world. Evans and Over speculated that System 2 operates on representations supplied by System 1 processes, and that these representations are then used to generate inferences and form judgments. Sloman as well as Stanovich and West described System 2 as having an interactive relationship with System 1. This is evident in instances in which both systems overlap, situations in which Sloman (1996) and Stanovich (1999; Stanovich & West, 2000) claimed there might be a conflict because the systems cue different responses. They suggested that, in the early tutoring studies, participants initially solved the task using System 1 processes. However, the initial matching response (PQ) conflicts with the logically valid response (P ¬Q), and the former response still influences participants' selection decisions because it is more salient.

Evans (1998) claimed that the selection task is a special case, in which heuristic processes determine card choices. In the arbitrary versions, the high proportion of PQ responses in the selection task has been interpreted as evidence of a matching bias: the tendency to consider cases as relevant when the lexical content of a case matches that of the conditional. When presented with a conditional statement that includes negations (e.g., "If there is no vowel on one side, then there is no even number on the other"), people show great difficulty in interpreting the statement and reasoning from it (see Evans, 1998). People tend to select the named cards and ignore the presence of the negations: in this case, selecting the vowel and even number. Evans (1998) claimed that only cards that appear to be relevant are selected, using heuristics, which are triggered automatically by surface level cues that are either linguistic or pragmatic. Such cues can focus attention on irrelevant rather than logically relevant information. Discuss-

ing early tutoring studies and Cheng et al.'s (1986) study, Evans and Over (1996) claimed that these other findings support their dual system account of reasoning. In their view, because the selection task is solved exclusively through System 1 processes, it is not surprising that tutoring techniques designed to engage System 2 processes fail to improve performance, because people cannot go beyond their initial biased judgments of relevance.

Since the development of the dual system framework of reasoning, various experimental techniques have been used to examine the contribution of analytic and heuristic processes to the selection task. For example, Evans, Ball, and Brooks (1987)—in a study designed to measure attentional biases by examining the order in which select/nonselect decisions were made for each card—claimed that people decide which cards to select before deciding which to reject (nonselections), because preconscious processes first focus attention on cards determined to be relevant, rather than on cards determined to be nonrelevant. Evans et al. (1987) found that matched cards were the most frequently selected, and that selections were made earlier than nonselections. A recent eye-tracking study (Ball, Lucas, Miles, & Gale, 2003) designed to measure attentional processing, and Evans's (1996) inspection time study, corroborated the original findings by Evans et al. (1987).

Roberts and Newton (2001) designed the rapid response time task to obtain purer measures of the dominant role of heuristic processes in reasoners' card responses. The example cards were presented one at a time, for 1 sec each, and participants were allowed 1 sec more to decide whether to select or reject the cards. These manipulations were designed to curtail analytic processes and to encourage participants to rely on heuristic processes. Roberts and Newton found that responses to the rapid response task did not radically differ from responses to free time versions (unlimited response time). However, they also found that, compared with the free time condition, there was an increase in the selection of the matching consequent card (Q) in the rapid response task. The fact that there were more matching consequent choices in the rapid response condition, which should reflect primary heuristic processing, than in the free time condition suggests that people in the free time condition were able to override the matching bias, presumably as the result of using analytic processes (Green, 1995a).

Deadline and inspection time procedures have been used to examine claims made by dual system theories. However, this can lead to problems, because dual system theories of reasoning do not identify the terms *heuristic processing* and *analytic processing* with particular experimental manipulations or data structures. For example, if tutoring is found to improve performance on the selection task when presented in both deadline and free time formats, then it is necessary to identify which process, heuristic or analytic, is being manipulated through tutoring. In the dual system framework, heuristic processing determines which cards are relevant, and these processes are inaccessible to the reasoner. Therefore, improvements in performance should result from tutoring modulating analytic processes, which are used to override such initial erroneous tendencies as matching bias. However, heuristic

processing determines what cases are relevant, and tutoring is one way of influencing what aspects of the task are considered to be so. Therefore, tutoring could modulate heuristic processing and its outputs. Both descriptions are equally plausible, but there is no data structure that could distinguish between them.

The present study does not make any claims about the underlying structure of the reasoning system, and therefore the terms *heuristic* and *analytic* are not assumed to be characteristic examples of functionally distinct reasoning systems. In the present study, heuristic processing is identified as understanding the selection task, and the different understandings of the task are evident in the range of card combinations selected (Gebauer & Laming, 1997; Osman & Laming, 2001). Through analytic processing—the evaluation of the statement and the cards—individuals can correct their misunderstandings, as evidenced by improvements in performance (Stenning & van Lambalgen, 2001, 2004). The simple account proposed here is that people make erroneous card choices in free time and deadline conditions, and that these are the result of misunderstandings of the statement or the cards or both. If the tutoring technique used in the present study is successful, it should improve performance, whether or not the transfer task is presented in deadline or free time format.

In order to examine improvements in performance on the selection task, the present study includes both deadline procedures (conditions in which time restrictions are imposed) and free time procedures. The innovation in the present study is the use of the deadline paradigm in combination with previously successful tutoring techniques (e.g., Socratic tutoring methods), to examine relative improvements in performance. Thus far, to the author's knowledge, no empirical work has compared performance in the selection task before and after tutoring, under both deadline and free time conditions. Therefore, the main objective of the present study is to examine whether tutoring methods designed to identify and address participants' misunderstandings of the task improve performance on tasks presented under deadline conditions.

EXPERIMENT 1

The facilitation of correct performance through tutoring was examined in a free time version of the selection task, and in two deadline versions of it: rapid response and rapid presentation time conditions. In the rapid response condition, cards were flashed for 1 sec and participants were allowed only 1 sec to respond. In the rapid presentation condition, cards were flashed for 100 msec, but participants were given 5 sec to respond. As described above, Roberts and Newton (2001) found that the rapid response condition produced an increase in selection of the matching consequent card, presumably as a result of heuristic processing. Heuristic processing should be encouraged by the even shorter displays in the rapid presentation condition, but the additional response time (RT) may allow some analytical processing to occur. Dominowski (1995) reported that the average RTs for P, Q, and ¬P cards ranged between 5.5 sec and 7.2 sec, whereas

for the ¬Q card the average RT was 13 sec. The 5-sec RT is the lower end of the range reported by Dominowski, particularly for decisions concerning the ¬Q, which is critical to performing the selection task correctly. Performance should improve under all three conditions if tutoring techniques designed to identify and reduce misunderstandings in arbitrary versions of the selection task are successful.

Method

Participants. Ninety undergraduate and graduate students from University College London were each paid £3 for taking part in the experiment. They were randomly assigned to each of the three conditions. Participants were screened for previous knowledge of the task. "Time outs" were occasions in which participants did not make a decision (i.e., a selection or nonselection) within the allotted time. Participants who timed out for half or more of the total responses within a set of problems (i.e., 12 total responses) were replaced. Three participants were replaced in the rapid response condition and two participants in the rapid presentation condition. In all other cases, time outs were treated as nonresponses. Participants were tested individually in sound-dampened testing cubicles, and the experimenter was present after the first session, in which participants received tutoring.

Materials. All versions of the selection task presented in Experiment 1 were arbitrary nonnegated problems. Two sets of problems were used: Set A and Set B. They are presented in Table 1.

Set A included three problems that were always presented prior to tutoring. Set B included three different problems that were presented directly after the tutoring session. For each participant, the order of presentation of the three problems from each set was randomized, as was the presentation of the four example cards for each problem.

Apparatus. All problems, with the exception of the one presented in the tutoring session, were presented on PCs with 43-cm color monitors and standard QWERTY keyboards. Responses were mouse controlled. Response times for selections and nonselections were recorded. Participants recorded their confidence ratings for each decision by typing a number corresponding to the appropriate point on the scale 1 (*not very confident*) to 7 (*highly confident*).

Procedure. The instructions presented on the screen informed participants that they would be solving three problems, each consisting of a statement and four cards, and that they would be shown only one side of each card. They were told that, after they had read each statement, the four cards would be shown one at a time, and that, in each case, they were to decide which card they would turn over. To indicate whether they would like to turn the card over, they were instructed to use the mouse cursor and press the *Yes* or *No* button to indicate that they did or did not wish to turn the card over. Participants then proceeded to the next screen, in the center of which was a blank rectangle, 4 × 6 cm. Directly below this were two mouse operated response buttons labeled Yes and No. At the bottom right of

the screen was a start button, which disappeared when pressed. After indicating that they had understood the instructions, participants began the free time condition by completing a practice problem, which was presented before the three problems from Set A.

At the start of each problem, the onscreen information included a conditional statement, which remained at the top of the screen until the problem had been solved. Participants began each problem by pressing the start button. This activated the two response buttons, the response timer, and the presentation of the example card, which appeared in the center of the screen, in place of the blank black rectangle. Participants were free to view the card for as long as they wished, and were under no time restrictions in making a decision. A decision was categorized as a selection if the Yes button was pressed, and categorized as a nonselection if the No button was pressed. After a decision had been made, a confidence rating scale (1–7) and a small text box appeared at the bottom of the screen. To rate their confidence in the decision they had just made, participants entered a number in the text box provided. The scale and text box then disappeared, and the start button reappeared on screen, to prepare the participant for the next card presentation. This procedure continued until all four example cards had been presented. After the confidence judgment for the fourth and final card of the problem, the start button initiated the presentation of a different conditional statement, and the procedure for card presentations, responses, and confidence ratings was repeated. After all three problems from Set A had been solved, the experimenter began the tutoring session.

The tutoring session used a Socratic teaching method, in combination with other techniques that have been found to successfully facilitate correct performance (e.g., Green, 1995b; Green & Larking, 1995; Platt & Griggs, 1995; Stenning & van Lambalgen, 2001). Participants were presented with a paper-and-pencil version of one of the problems presented in Set A, and they were asked to think aloud while solving the task. This procedure enabled the experimenter to identify the kinds of misinterpretation participants had of the conditional statement, and to determine which cards they considered relevant to the task. Participants were then told to imagine the possibilities on the underside of each of the four cards, and to think about what bearing each possibility would have on the validity of the statement. They were then asked to name cards that would violate the statement. A short tutorial about the meaning of falsification was given, along with an explanation of the different effects that different cards had on the statement's truth value. After tutoring, participants returned to the computer to complete the second session, during which problems from Set B were presented, using the same procedure used to present problems from Set A.

The procedure described here was also adopted in the rapid response and rapid presentation conditions, except that the presentation times of the cards were varied and response deadlines were introduced.

In the rapid response task, the free time and rapid response conditions differed in three respects. The instructions included the fol-

Table 1
Statement and Example Cards Used in Problem Sets A and B

Set	Statement	Example Cards
A	If there is an A on one side of the card, then there is a 2 on the other side.	A, G, 2, 7
	If there is a square on one side of the card, then there is a red background on the other side.	Square, Diamond, Red card, Green card
	If there is picture of a table on one side, then there is picture of a dog on the other side.	Table, Chair, Dog, Cat
B	If there is a K on one side of the card, then there is a £ sign on the other side.	K, E, £, $
	If there is a triangle on one side of the card, then there is a yellow background on the other side.	Triangle, Square, Yellow card, Blue card
	If there is picture of a hat on one side, then there is picture of a car on the other side.	Bowler Hat, Stiletto shoe, Car, Bicycle

lowing sentence: "Each card you are shown will be flashed for one second, and you will have a further second to decide whether you need to turn it over or not in order to test the statement." For each problem, each card was then flashed for 1 sec. The two response buttons and the response timer were then activated. Timing stopped when a decision had been made, or when 1 sec had elapsed.

In the rapid presentation task, the presentation time of the example cards was reduced from 1 sec to 100 msec, and the response deadline was increased from 1 sec to 5 sec. The instructions included the following sentence: "Each card you are shown will be flashed for a brief period, and you will have a further 5 sec to decide whether you need to turn it over or not in order to test the statement." In all other respects, the procedure was identical to that used in the free time and rapid response conditions.

Results

This section begins with an analysis of card selection behavior, and then examines the effects of tutoring on confidence ratings. In all analyses reported in this section and for Experiment 2, a significance criterion of $\alpha = .05$ was used.

Card choices. For each participant, a percentage was calculated for each card, on the basis of the card's selection within a problem set. Depending on the percentage calculated for each card, a second scoring scheme was used on the basis of Roberts's (1998) scoring scheme: logical antecedent choices, calculated from the percentage of the correct antecedent card (P) selected and the nonlogical antecedent card (¬P) not selected; and logical consequent choices, calculated from the percentage of the logical consequent card (¬Q) selected and the nonlogical consequent card (Q) not selected. For example, if a participant selected the P card on all three problems and the ¬P card once, they would be coded as P card = 100% and ¬P = 33%, and their actual logical antecedent choices score would be 83.5%: [% P selections (e.g., 100) + % ¬P nonselections (e.g., 67)] / 200 = 0.835 ∗ 100. The same rationale was used to calculate logical consequent choices, except that the scoring was based only on the consequent cards. Both procedures were carried out separately for each participant. Table 2 presents the mean logical antecedent choices and logical consequent choices.

For the card selection data, Table 2 presents the mean percentages for each card choice in each condition, before tutoring (Session 1) and after tutoring (Session 2). The trends suggested by these figures are more clearly indicated in the rightmost columns of Table 2, labeled *Logical Antecedent* and *Logical Consequent*. Looking down the column Logical Antecedent, the scores appear to be stable across conditions in Session 1 and Session 2. For logical consequent choices, the trend indicated by these scores is in the expected direction, for both deadline conditions in Session 1. The scores indicate, consistent with Roberts and Newton's (2001) findings, that in deadline conditions participants made fewer logical consequent choices than in free time conditions. Table 2 also indicates that logical consequent choices increased in all three conditions after tutoring.

A 2 × 3 mixed design ANOVA was used to analyze logical antecedent choices, with tutoring (Sessions 1 and 2) as the within-subjects variable, and condition (rapid response, rapid presentation, and free time) as the between-subjects variable. This analysis revealed no significant main effect of tutoring [$F(1,87) = 0.04$, $MS_e = 14.7$], no significant main effect of condition [$F(2,87) = 0.11$, $MS_e = 60.6$], and no significant tutoring × condition interaction [$F(2,87) = 0.50$, $MS_e = 200.0$]. A parallel analysis conducted of logical consequent scores revealed a significant main effect of tutoring [$F(1,87) = 35.56$, $MS_e = 21,659.2$]. There was a main effect of condition [$F(2,87) = 4.23$, $MS_e = 3,100.3$]. The tutoring × condition interaction approached significance [$F(2,87) = 3.07$, $MS_e = 1,869.8$, $p = .052$]. Paired sample t tests among the three condition means revealed that in Session 2, logical consequent choices for the free time condition differed significantly from the rapid presentation condition [$t(29) = 2.08$, $SE = 8.36$]. In Session 2, logical consequent choices in the free time condition differed significantly from the rapid response condition [$t(29) = 2.99$, $SE = 7.74$]. No other analyses were significant.

Thus, performance improved after tutoring, as indicated by the increase in logical consequent choices in Session 2. Overall, the change in logical consequent scores after tu-

Table 2
**Mean Percentage for Each Card Selected and Logical Antecedent and Consequent Choices
in Each Condition for Both Sessions in Experiments 1 and 2**

			P		¬P		Q		¬Q		Logical Antecedent		Logical Consequent	
	Condition	Session	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Experiment 1	Free time	1	83	39	21	41	46	48	17	38	85	25	37	28
		2	93	25	17	38	37	49	70	47	88	25	66	33
	Rapid response	1	86	19	15	24	54	28	20	22	86	17	32	18
		2	82	27	10	20	34	38	23	29	84	15	45	27
	Rapid presentation	1	85	23	14	30	54	27	16	27	86	22	32	18
		2	86	27	15	29	31	34	29	35	82	24	55	28
Experiment 2	No tutoring, rapid	1	81	31	24	34	43	37	12	24	80	24	35	22
	presentation	2	79	28	22	36	50	40	18	32	80	21	33	26
	Tutoring, rapid	1	81	30	23	35	45	39	14	26	82	25	35	22
	presentation	2	83	23	15	27	34	39	29	34	83	18	47	30

Table 3
Overall Mean Confidence Ratings (Scale, 1–7), by Condition,
for Experiments 1 and 2

			Selections		Nonselections	
	Condition	Session	M	SD	M	SD
Experiment 1	Free time	1	3.4	1.1	3.4	1.3
		2	4.3	1.9	4.4	1.7
	Rapid response	1	2.7	1.6	3.6	1.6
		2	3.4	1.8	3.9	1.6
	Rapid presentation	1	3.2	1.0	3.0	1.3
		2	4.2	1.1	4.2	1.2
Experiment 2	No tutoring, rapid	1	3.2	1.1	2.7	1.0
	presentation	2	4.5	1.1	4.3	1.3
	Tutoring, rapid	1	3.2	1.4	3.5	1.0
	presentation	2	4.6	1.4	3.8	1.7

toring is primarily the result of a decrease in Q cards rather than an increase in ¬Q card choices.

Confidence ratings. Participants' confidence ratings were computed separately for selections and nonselections. The overall mean ratings for selections (cards chosen) and nonselections (cards rejected) are presented in Columns 4 and 6 of Table 3. The figures suggest that, after tutoring, participants in deadline and free time conditions had more confidence in their decisions. This was confirmed using a $2 \times 2 \times 3$ mixed design ANOVA with tutoring (Sessions 1 and 2) and decision (selection, nonselection) as within-subjects variables; and condition (rapid response, rapid presentation, free time) as a between-subjects variable. There was a significant main effect on ratings of tutoring [$F(1,87) = 37.44$, $MS_e = 68.5$]. There was also a main effect of condition [$F(2,87) = 3.29$, $MS_e = 7.8$]. There was no main effect of decision [$F(1,87) = 1.79$, $MS_e = 4.2$] and no significant interactions ($F < 1$). Paired sample t tests among the three condition means revealed that, in Session 2, confidence ratings for the rapid response condition differed significantly from the free time condition [$t(59) = 2.48$, $SE = 0.20$]. No other analyses were significant. Thus, in the rapid response condition, participants were less confident in their decisions than in the free time condition.

Further analyses were conducted to examine whether confidence ratings corresponded with performance. Pearson's correlation coefficient did not reveal significant correlations when confidence ratings were differentially correlated with logical antecedent scores and logical consequent scores for the three different conditions by session. However, confidence ratings correlated significantly with RTs. Correlations were carried out separately, for each condition based on ratings and RTs, for selection responses in Session 1 and then for Session 2. The same analyses were then carried out, based on confidence ratings and RTs, for nonselections in Session 1 and Session 2. Significant negative correlations between ratings and RTs were found in each of the three conditions in Session 2, based on selections only: rapid presentation condition [$r(30) = -.46$, $p < .05$], rapid response condition [$r(30) = -.40$, $p < .05$], free time condition [$r(30) = -.58$, $p < .05$]. No other correlations were significant. Thus, after tutoring, participants' confidence ratings cor-

responded to the speed at which they made a selection: The faster participants' responses, the higher their confidence in their decisions.

Discussion

The evidence from Experiment 1 can be summarized as follows: First, the Socratic teaching method, in combination with the other tutoring techniques, improved performance in deadline conditions and in the free time equivalent. In particular, tutoring increased logical consequent scores in all three conditions, whereas logical antecedent scores remained stable across conditions. In arbitrary versions of the task, such as those used previously, the main misunderstanding results from the consequent cards, as is evident in the failure to select the correct card, ¬Q, and selecting the Q card instead. In contrast, the typical selection behavior for antecedent cards is to select the correct card, P, and rarely to make the error of selecting ¬P. Thus, in the present study, the proportion of logical antecedent choices remained stable in both sessions. The tutoring that participants received between sessions merely reinforced their decisions for antecedent cards. Therefore, the success of the Socratic tutoring method was observed in the elevation of logical consequent scores, which reflects the correct selection of the ¬Q card in conjunction with the rejection of the Q card.

Second, the rapid presentation condition was designed as a deadline condition different from the rapid response condition, and the manipulations in the speed of presentation of the example cards were designed to elicit card choice patterns similar to the rapid response task. There was a trend in the expected direction for both deadline conditions, as indicated by the depressed performance in Session 1, compared with the free time condition. Roberts and Newton (2001) reported that curtailing thinking times in the rapid response condition altered responses. Participants made more matching consequent choices in the rapid response condition than in the free time condition.

Third, after tutoring in all three conditions, participants had higher ratings of confidence in their decisions. Despite this, overall confidence judgments for decisions in the rapid response condition were lower than in the free time condition. One reason for this result may be the short RT within which participants had to make a decision. In

addition, confidence judgments of selection decision corresponded to the speed with which that decision was made in Session 2. There is evidence to suggest that the fluency of processing (the ease with which stimuli are perceived or manipulated) is used as a basis for making judgments (e.g., Johnston, Hawley, & Elliott, 1991; Mandler, 1980; Verfaellie & Cermak, 1999; Whittlesea, 1993). In the present study, participants may have used the fluency with which they made selection decisions as an indication of the accuracy of their choice, and so judged faster selections more confidently, regardless of the actual accuracy of that decision.

Why was tutoring successful in Experiment 1? The facilitatory effects of tutoring may have been strengthened because the transfer tests directly followed the tutoring session. Therefore, the information with which participants were presented may have remained fresh in their minds when they came to perform the tasks for themselves, and this contributed to the dramatic increases in performance across conditions. To establish the robustness of the tutoring technique, a second experiment was therefore run. Furthermore, there was no control condition against which to compare the elevation in confidence judgments after tutoring. In Experiment 2, therefore, a no-tutoring control condition was used.

EXPERIMENT 2

Experiment 2 was designed with two purposes in mind. First, given that the rapid presentation deadline procedure used in Experiment 1 had not previously been tested, a replication was necessary. Second, Experiment 1 successfully showed that the facilitatory effects of tutoring transferred to free time and deadline presentation formats. However, performance may have improved because transfer tests immediately followed the tutoring session, and because by the time they came to solve the second set of problems, participants were highly familiar with the task formats. Thus, Experiment 2 included two conditions: tutoring and no-tutoring. In the tutoring condition, participants solved problems from Set A presented under rapid presentation deadline conditions, and, after completion, they were tutored using the same instructions as those used in Experiment 1. After a day, participants returned to complete the remainder of the experiment and solved problems from Set B under the same test conditions as those for Set A. In the no-tutoring condition, exactly the same procedure was used as in the tutoring condition, except that participants received no tutoring after completing problems in the first test session.

Method

Participants. Sixty undergraduate and graduate students from University College London took part, 30 in each condition. Participants were each paid £3 on completion of the experiment, and all were screened for prior knowledge of the selection task. Experiment 2 used the same criteria as Experiment 1 to eliminate participants who "timed out." A total of 3 participants were replaced.

Materials and Apparatus. The same two sets of problems, Sets A and B, were used as in Experiment 1, along with the instructions for the rapid presentation condition. The experiment was fully automated, using the same apparatus as in Experiment 1.

Procedure. The procedures in the tutoring and no-tutoring conditions were exactly the same as in the rapid presentation condition in Experiment 1, with two exceptions. In the tutoring condition, participants solved problems from Set A, then received tutoring. The following day, they returned to solve problems from Set B. In the no-tutoring condition, participants did not receive tutoring between the presentation of the Set A and Set B problems. In both conditions, participants were unaware of what they were expected to do in the second session of the experiment. They were told only that they had to return after a day to complete the experiment.

Results

This section begins with an analysis of card selection behavior, then examines the effects of tutoring on confidence ratings. Card choices were scored in exactly the same way as in Experiment 1.

Card choices. Table 2 shows the mean proportion of selections of each card and logical antecedent and logical consequent choices in Sessions 1 and 2. Before tutoring, logical antecedent and logical consequent choices were similar in the tutoring and the no-tutoring conditions, but in the second session more logical consequent choices were made in the tutoring condition. To analyze this, a 2×2 mixed design ANOVA was conducted on logical antecedent scores with session (Sessions 1 and 2) as the within-subjects variable and tutoring (tutoring, no tutoring) as the between-subjects variable. This analysis revealed no significant main effect of session [$F(1,58) = 0.59$, $MS_e = 221.4$], no main effect of condition [$F(1,58) = 1.24$, $MS_e = 765.1$], and no session $\times$ tutoring interaction [$F(1,58) = 2.03$, $MS_e = 765.1$]. The same analysis was performed on logical consequent scores. This analysis revealed no significant main effect of session [$F(1,58) = 2.84$, $MS_e = 989.0$] and no main effect of tutoring [$F(1,58) = 1.55$, $MS_e = 1,404.3$]. However, there was a session $\times$ tutoring interaction [$F(1,58) = 4.05$, $MS_e = 1,411.1$]. Paired sample t tests revealed that the difference in logical consequent choices between Session 1 and Session 2 was significant in the tutoring condition [$t(29) = 2.35$, $SE = 5.4$], but was not significant in the no-tutoring condition [$t(29) = 0.27$, $SE = 4.2$].

Confidence ratings. Table 3 suggests that participants' confidence ratings increased in the second session of both conditions. A $2 \times 2 \times 2$ mixed design ANOVA on confidence ratings was carried out with session (Session 1, Session 2) and decision (selection, nonselection) as within-subjects variables, and tutoring (tutoring, no tutoring) as a between-subjects variable. There was a significant main effect of session [$F(1,58) = 40.93$, $MS_e = 74.8$] and a significant main effect of decision [$F(1,58) = 4.41$, $MS_e = 6.0$]. The analysis did not reveal a significant main effect of tutoring [$F(1,58) = 0.32$, $MS_e = 0.4$], and no significant interactions ($F < 1$). Overall, the analysis indicates that participants were more confident making selection decisions than making nonselection decisions. In addition, in the second session, experience with the deadline format increased participants' confidence in their decisions, regardless of whether they had received tutoring.

Consistent with Experiment 1, there were no significant correlations between confidence ratings and logical antecedent scores or logical consequent scores. However, confidence ratings correlated significantly with RTs. Significant negative correlations between ratings and RTs were found only in the tutoring condition in Session 2, based on selections only [$r(30) = -.52$]. No other correlations were significant. Thus, after tutoring, participants' confidence judgments corresponded to the speed of their selection decision: The faster they made their decision, the higher their confidence judgment.

Discussion

The evidence from Experiment 2 can be summarized as follows. First, because there was no-tutoring condition in Experiment 1, the elevated performance found in Session 2 may have been the result of repeated exposures to the task, or may have been facilitated by tutoring. In response to this, the evidence from Experiment 2 indicates that many exposures to the same task format do not of themselves improve performance in the selection task. Consistent with the findings from Experiment 1, there was an increase in logical consequent choices for the tutoring group in Session 2. Thus, the facilitatory effects of tutoring were transferred to problems presented under deadline conditions when an interval of a day separated tutoring and test.

Second, there was a general increase in confidence ratings in the second test session, regardless of condition. This result suggests that the high ratings of confidence in Session 2 in Experiment 1 may have been the result of practice and not tutoring. However, as with Experiment 1, confidence judgments for selections made in Session 2 corresponded with the speed with which selection decisions were made, and this result was present only in the tutoring condition. This result suggests that, for selection decisions, the basis on which confidence judgments were made differed between the tutoring and no-tutoring conditions. The indication that RTs corresponded with confidence rating suggests that in the tutoring condition, participants perceived the fluency of their decisions as a useful indicator of their performance, and therefore judged their choices as more confident. In contrast, in the no-tutoring condition, fluency was not the basis for making confidence judgments, since there could be no influence of tutoring to modify participants' perception of their responses. Instead, general familiarity with the task format may have formed the basis for their confidence judgments.

GENERAL DISCUSSION

The main finding from the two experiments in this study was that tutoring was successful in facilitating correct performance on free time and deadline versions of the selection task. In the present study, tutoring was designed to address misinterpretations of the statement (Gebauer & Laming, 1997; Osman & Laming, 2001) and errors in the values assigned to the underside of the cards. This is compatible with Almor and Sloman's (1996) claim that

performance on the selection task is determined by the expectations of what should be on the underside of each card. They predicted that the greater the expectation of the hidden value of the card, the more likely that card would be selected. Almor and Sloman (2000) studied this by making explicit the expectedness of the alternative outcomes, using nondeontic rules (e.g., if a product gets a prestigious prize, it must have a distinctive quality). They found that the frequency of P and ¬Q selections was equivalent to that reported by studies using deontic content (e.g., Cheng & Holyoak, 1985). The argument proposed here is that, in the present study, tutoring techniques emphasized the relevance of P and ¬Q cards, but also guided participants towards having expectations for the hidden values of these cards, and that this led to significant improvements in performance. More important, this was evident in deadline conditions, suggesting that even under speeded conditions participants are able to correct their initial misunderstandings of the task.

Evans proposed that preconscious processes determine participants' choices in the selection task. He also argued that examples in which people can adapt their reasoning processes through instruction imply the operation of an explicit thinking system that is under people's conscious control (Evans, 2000). This position has been supported by studies using the inspection-time paradigm (e.g., Ball et al., 2003; Evans, 1996) and deadline conditions (e.g., Schroyens, Schaeken, & Handley, 2003). In both of these task formats, preconscious processes are implicated in the card choices that are made. If this particular position is applied to the findings from the present study, then the evidence here suggests that in fact there is no need to posit any preconscious heuristic processes, and that instead participants are able to solve the task using entirely analytic processes. Given that participants are able to access and modify their card choices under deadline conditions after tutoring, it seems plausible that analytic processes operate more quickly than was initially claimed by those advancing dual system theories. However, as discussed in the introduction, problems arise from the fact that dual system theories of reasoning do not identify the terms *analytic* and *heuristic* with particular experimental manipulations or data structures. Therefore, the only foundation on which participants can be said to be using heuristic or analytic processes is based on the definitions of these terms outlined in the introduction. On this basis, the findings from Experiments 1 and 2 demonstrate that participants commonly begin the task with a misunderstanding of the statement and the cards, and that tutoring is one example of a corrective method that attenuates participants' misunderstandings. Beyond these tentative conclusions, there is no secure basis for implicating preconscious processes in deadline conditions, or for positing the kind of relationship shared between System 1 and System 2 processes.

The practice effects, in combination with tutoring, may have greatly elevated performance after tutoring, and the addition of a delay between tutoring and test in Experiment 2 was designed to examine this phenomenon. It could be argued that a day's delay was insufficient to examine the success of the tutoring technique used in the

present study. However, since the problems presented after tutoring in Experiment 2 were under deadline conditions, in which the task is harder to solve, this delay was adequate to test the success of tutoring.

To the author's knowledge, there has been no empirical work on the selection task that includes confidence measures in a deadline format. Because of the exploratory nature of the study, any conclusions are to be drawn with caution. One of the consistent findings from Experiments 1 and 2 was that participants' confidence ratings increased in the second session. However, the inclusion of the control condition in Experiment 2 clarified the fact that the overall increase in confidence ratings in Experiment 1 was the result of practice, not of the tutoring participants received between sessions. Experiment 1 also indicated that, overall, participants were less confident in their decisions in the rapid response condition than in the free time condition. One reason for this is simply that participants had less time to respond in the rapid response condition than in the free time condition, and that this affected their confidence, because they felt less secure about their decisions. In addition, if confidence is related to analytical processes, the lower confidence ratings found in the rapid response condition were to be expected.

The second main finding was that all tutoring conditions included in the study showed an inverse relationship between confidence ratings and RTs: Confidence ratings increased as RTs decreased for selection decisions in the second session. This pattern is comparable to that of the fluency effect, in which processing of any stimulus can vary across different parameters nonspecific to the content of the stimulus. For instance, the speed rather than the accuracy of stimulus processing, or "processing fluency," is used as a basis for making a host of judgments, and is considered to be a result of priming procedures. To illustrate, Reber and Schwarz (2001) reported that the same targets were judged more favorably when preceded by matching rather than by nonmatching prime targets. In a similar vein, the correlation between confidence ratings and RTs found in the second session reflects the fact that speed of processes, rather than accuracy, may have been used as a guide for confidence ratings.

In sum, the Socratic tutoring procedure that was used in the study was based on identifying and modifying the kinds of misunderstandings participants had of the task. This type of tutoring helped to create specific memories related to the task, and when the task was next encountered, these memories were recalled to shape participants' understanding. The effectiveness of tutoring was achieved through a combination of factors that include addressing participants' particular misunderstandings. In addition, the success of tutoring depends on the particular relationship between those memories formed during tutoring and the version of the task presented to participants. The present study demonstrates that effective tutoring can be achieved when available cognitive resources are reduced, either by imposing a deadline for making decisions (i.e., rapid response task), or by examining stimuli (i.e., rapid presentation task). Further empirical work exploring dif-

ferent types of deadline conditions may offer a way of better understanding the capabilities of the reasoning system, but the present study suggests that, currently, dual system theorists have misclassified, as evidence of implicit processes, the reasoning phenomena that have been studied using deadline and inspection time procedures. As yet, the precision of the theoretical accounts (for a discussion, see Osman, 2004) and the methodology are not sufficient to make these claims.

REFERENCES

ALMOR, A., & SLOMAN, S. (1996). Is deontic reasoning special? *Psychological Review*, **103**, 374-380.

ALMOR, A., & SLOMAN, S. (2000). Reasoning versus text processing in the Wason selection task: A nondeontic perspective on perspective effects. *Memory & Cognition*, **28**, 1060-1070.

BALL, L. J., LUCAS, E. J., MILES, J. N. V., & GALE, A. G. (2003). Inspection times and the selection task: What do eye-movements reveal about relevance effects? *Quarterly Journal of Experimental Psychology*, **56A**, 1053-1077.

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 tutoring 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.

DOMINOWSKI, R. L. (1995). Content effects in Wason's selection task. In S. E. Newstead & J. S. B. T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honor of Peter Wason* (pp. 41-65). Hove, U.K.: Erlbaum.

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

EVANS, J. S. B. T. (1998). Matching bias in conditional reasoning: Do we understand it after 25 years? *Thinking & Reasoning*, **4**, 45-82.

EVANS, J. S. B. T. (2000). What could and could not be a strategy in reasoning. In W. Schaeken, G. De Vooght, A. Vandierendonck, & G. d'Ydewalle (Eds.), *Deductive reasoning and strategies* (pp. 1-22). Mahwah, NJ: Erlbaum.

EVANS, J. S. B. T., BALL, L. J., & BROOKS, P. G. (1987). Attention bias and decision order in a reasoning task. *British Journal of Psychology*, **78**, 385-394.

EVANS, J. S. B. T., & OVER, D. E. (1996). *Rationality and reasoning.* Hove, U.K.: Psychology Press.

GEBAUER, G., & LAMING, D. (1997). Rational choices in Wason's selection task. *Psychological Research*, **60**, 284-293.

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

GREEN, D. W. (1995a). The abstract selection task: Thesis, antithesis, and synthesis. In S. E. Newstead & J. S. B. T. Evans (Eds.), *Perspectives on thinking and reasoning* (pp. 173-188). Hove, U.K.: Erlbaum.

GREEN, D. W. (1995b). Externalisation, counter-examples and the abstract selection task. *Quarterly Journal of Experimental Psychology*, **48A**, 547-574.

Green, D. W., & Larking, R. (1995). The locus of facilitation in the abstract selection task. *Thinking & Reasoning*, **1**, 183-199.

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

Hoch, S. P., & Tschirgi, J. E. (1983). Cue redundancy and extra logical inferences in a deductive reasoning task. *Memory & Cognition*, **11**, 200-209.

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

Johnston, W. A., Hawley, K. J., & Elliott, J. M. (1991). Contribution of perceptual fluency to recognition judgments. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **17**, 210-223.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, **87**, 252-271.

Margolis, H. (2000). Wason's selection task with a reduced array. *Psycoloquy*, **11**(5).

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

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

Ormerod, T. C., Manktelow, K. I., & Jones, G. V. (1993). Reasoning with three types of conditional: Biases and mental models. *Quarterly Journal of Experimental Psychology*, **46A**, 653-677.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, **11**, 998-1010.

Osman, M., & Laming, D. (2001). Misinterpretation of conditional statements in Wason's selection task. *Psychological Research*, **65**, 128-144.

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

Reber, R., & Schwarz, N. (2001). The hot fringes of consciousness: Perceptual fluency and affect. *Consciousness & Emotion*, **2**, 223-231.

Roberts, M. J. (1998). Inspection times and the selection task: Are they relevant? *Quarterly Journal of Experimental Psychology*, **51A**, 781-810.

Roberts, M. J., & Newton, E. J. (2001). Inspection times, the change task, and the rapid-response selection task. *Quarterly Journal of Experimental Psychology*, **54A**, 1031-1048.

Schroyens, W., Schaeken, W., & Handley, S. (2003). In search of counter-examples: Deductive rationality in human reasoning. *Quarterly Journal of Experimental Psychology*, **56A**, 1129-1145.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, **119**, 3-22.

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

Stenning, K., & van Lambalgen, M. (1999). Is psychology hard or impossible? Reflections on the conditional. In J. Gerbrandy, M. Marx, M. de Rijke, & Y. Venema (Eds.), *Liber amicorum for Johan van Bentham's 50th birthday* (pp. 1-29). Amsterdam: Amsterdam University Press.

Stenning, K., & van Lambalgen, M. (2001). Semantics as a foundation for psychology: A case study of Wason's selection task. *Journal of Logic, Language, & Information*, **10**, 273-317.

Stenning, K., & van Lambalgen, M. (2004). A little logic goes a long way: Basing experiment on semantic theory in the cognitive science of conditional reasoning. *Cognitive Science*, **28**, 481-529.

Verfaellie, M., & Cermak, L. (1999). Perceptual fluency as a cue for recognition judgments in amnesia. *Neuropsychology*, **13**, 198-205.

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

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

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

Wason, P. C., & Golding, E. (1974). The language of inconsistency. *British Journal of Psychology*, **65**, 537-546.

Wason, P. C., & Johnson-Laird, P. N. (1970). A conflict between selecting and evaluating information in an inferential task. *British Journal of Psychology*, **61**, 509-515.

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

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1235-1253.