Logical knowledge and cue redundancy in deductive reasoning

STEPHEN J. HOCH

University of Chicago, Chicago, Illinois

and

JUDITH E. TSCHIRGI

AT&T Bell Laboratories, Naperville, Illinois

This study clarified the basis of Hoch and Tschirgi's (1983) finding of good performance on abstract deductive reasoning problems based on Wason's (1966) four-card selection task. Different versions of the problem were developed through the addition of cues about antecedent-consequent relations redundant with the logical structure of the task. In all versions of the task, abstract symbols were used, and subjects were asked to test the truth or falsity of an implication rule. Twenty-five subjects from each of three education levels (high school, bachelor's, and master's) solved one of four versions of the task. Forty-eight percent of the master's subjects solved the original abstract version of the task compared to less than 10% of the high school and bachelor's subjects. When the problems contained redundant cues, performance improved dramatically, most notably in the bachelor's subjects. Reasoning performance seemed to be a function of both the general inferential abilities that subjects brought to the task and the redundant cues conveyed through problem content. Many of the master's subjects had adequate logical knowledge to solve the problems without the addition of redundant cues, though performance did increase when such cues were present. In contrast, high school subjects appeared to have so little understanding of the logical structure of conditionals that cue redundancy improved performance only slightly.

In his 1952 monograph, "The Conceptual Framework of Psychology," Brunswik argued that human performance is determined by the perception and integration of multiple probabilistic cues. Typically, individual cues in the environment provide only partially valid information about the objects of perception. Performance, however, is a function not only of the individual cues but also of the redundancy between those cues (i.e., inter-cue correlations or vicarious functioning) and the ability of a perceiver to detect and integrate those cues. Although performance may be low when a subject has access to only one imperfect cue, performance improves when the subject can rely on a number of imperfect but correlated cues. Deductive reasoning tasks, such as the Wason (1966) four-card selection task, can be analyzed within this framework (Hoch & Tschirgi, 1983). The problem content consists of various cues to an underlying structure. The goal of the problem solver is to perceive the valid cues to logical structure and infer a solution. However, people vary in their abilities to perceive cues and in their own inferential machinery. In this paper, we examine the selection task in terms of cue redundancy and differences in general inferential ability among members of the subject population.

The four-card selection task is one of the more widely researched deductive reasoning tasks, primarily because of the surprising nature of the results (Wason, 1983). The task is deceptively simple. The subject is given a rule, "If a card has a vowel on its letter side, then it has an even number on its number side," and is shown four cards face up: A, K, 18, 5. The subject has to decide which card(s) must be turned over to prove the truth or falsity of the rule. This task corresponds to the material-implication rule of the form, "If p then q ." The cards A and K represent the antecedents (p and $\neg p$, respectively), and the cards 18 and 5 represent the consequents (q and $\neg q$). The solution is to turn over A and 5 (p and $\neg q$), because the rule is violated only by cards pairing a vowel with an odd number [i.e., A, 5 ($p, \neg q$) and 5, A ($\neg q, p$)]. Most studies have found that only about 10% of the subjects can solve abstract forms of the problem (see Evans, 1982, for a review). Subjects usually have been undergraduates, but other researchers have reported that even more sophisticated subjects (e.g., PhD psychologists and statisticians) have difficulty with the task (Dawes, 1975; Einhorn & Hogarth, 1978; Griggs & Ransdell, 1985).

Since 1966, performance has improved in some studies using problems that are reformulated with thematic or realistic materials in lieu of the arbitrary, abstract symbols originally used (Evans, 1982). Thematic materials, however, have not always proved to be an advantage (Griggs & Cox, 1982). Simply replacing abstract sym-

Requests for reprints should be sent to Stephen J. Hoch, University of Chicago, Graduate School of Business, Center for Decision Research, 1101 E. 58th Street, Chicago, IL 60637.

bols with thematic items may not improve performance, if the antecedents and consequents remain arbitrarily related, as in the original rule (van Duyne, 1976). Using a problem based on "drinking age regulations" in Florida, Griggs and Cox (1982; Cox & Griggs, 1982) identified two mechanisms by which thematic materials could improve performance: memory cuing and reasoning by analogy. They concluded that subjects were using past experience with the specific rule to generate and recognize the important $(p, \neg q)$ counterexample. Facilitation occurred because of a coincidental congruence between pragmatic constraints dictated by past experience and the logical structure of material implication.

CUE REDUNDANCY AND EXTRA-LOGICAL INFERENCES

In Hoch and Tschirgi (1983) and in this study, reasoning tasks are viewed as containing at least two types of cues to underlying structure: (1) the implication rule, and all that it implies about that structure; and (2) relationships between antecedents and consequents revealed through problem content (e.g., cuing knowledge about real-world regulations). In abstract versions of the selection task, typically the only cue for solving the task is the logical rule. Theoretically, the implication rule provides enough information to construct a formal truth table and then to solve the problem. Most people, however, have only partial knowledge about the propositional calculus, and the logical cue by itself does not provide them with adequate information to solve the problem. When problem content provides information that coincides with the structure of the truth table, however, reasoning performance can improve (Hoch & Tschirgi, 1983). If deductive reasoning is viewed as one of a larger class of psychological tasks that requires perception of some underlying structure, then redundancy among cues can only improve this probabilistic process (Brunswik, 1952; Hammond, 1966).

Hoch and Tschirgi (1983) identified an additional strategy for solving this task not requiring subjects to invoke domain-specific, personally experienced knowledge to construct counterexamples or reason by analogy. First, subjects generate all of the possible antecedent-consequent pairs plus the four reverse orderings. Next, subjects identify the truth value of each pair [(p, q) , $(\neg p, q)$, $(\neg p, \neg q)$ are true and $(p, \neg q)$ is false] and the reverse orderings. These two steps represent the psychological analogue of the construction of a truth table. Finally, subjects select the two cards representing the false pair. Evidence from Johnson-Laird and Tagart (1969) and Wason and Johnson-Laird (1972) suggests that subjects seldom generate all four pairs in this task and often fail to recognize the reversibility of a single-ordered relation. We reasoned that, if subjects could overcome the generation obstacle, their performance might improve. Toward this end, we constructed problems with two different sets of cues to encourage subjects to enumerate all antecedent-consequent pairs. The cues either explicitly stated or merely implied

that the $\neg p$ antecedent could be validly paired with either q or $\neg q$ (or equivalently that the q consequent could be validly paired with either the p or $\neg p$ antecedents). Although the explicit problem provided extra-logical information, in no case were the subjects directly told the truth values of the (p, q) and $(p, \neg q)$ pairs. Using this information that a person familiar with material implication or with access to a truth table would know, subjects would have good reason not to turn over q , the most common error.

In both abstract (letters and numbers) and thematic (quality control scenario) problems, with either implicit or explicit relations, performance increased substantially over conditions with no relation between the antecedents and consequents. Moreover, an analysis of subjects' reasons for selecting and not selecting cards showed that subjects given implicit cues inferred the same extra-logical relations that the explicit cues provided directly. This is not to say that memory cuing and reasoning by analogy do not occur, even in abstract problems. Performance also improved when the $\neg q$ card was represented by a "blank" (e.g., a blank card in the abstract task). Subjects relied on general knowledge that the "absence" of something often is not normal, selecting the $\neg q$ card to verify what was on the other side.

The present experiments were conducted to clarify two issues. First, are the cue redundancy effects a stable finding? To our knowledge, no other published studies have found improved performance on problems using abstract materials. Hoch and Tschirgi's (1983) study contained several procedural differences that could limit generalizability. Instead of the true-false wording used in the original problem, subjects were asked to make sure that the rule was followed (see also Griggs & Cox, 1982). Yachanin and Tweney (1982) argued that the true-false version of the task is more difficult and requires fundamentally different psychological processes; two hypotheses need to be processed rather than only the one in the modified version. In the present study we returned to the true-false wording of Wason's original task (also see Griggs, 1984). Another modification introduced in the Hoch and Tschirgi (1983) study was to have subjects explain their selection decisions. Although subjects wrote these explanations after answering the problem, they were not prevented from changing their selections while generating their explanations. It is possible that explanations had a positive effect on reasoning performance, if for no other reason than that subjects had to spend additional time thinking about the problem (cf. Evans & Wason, 1976). In the present experiments, subjects were not allowed to change their initial responses; however, they were permitted to provide a second answer later, allowing us to monitor first responses and also to measure change arising from the explanation task.

The second issue concerns the generally high performance level (48%) found across the four different abstract problems in the Hoch and Tschirgi (1983) study, especially the 28% solution rate (compared to the usual 10%)

on the standard version that contained no cues beyond the implication rule. What accounted for the improved performance compared to past studies? The improved performance could be due to the procedural differences just cited. A second hypothesis is that the subject population, first- and second-year graduate students, differed from the more typical undergraduate populations of previous studies. The task requires matrix multiplication, analogical reasoning, and/or deductive reasoning, all hallmarks of formal operational thought (Inhelder & Piaget, 1958). Because all subjects in this task were young adults or older, maturational differences cannot account for the performance differences. Differences in reasoning expertise, however, may distinguish the different populations. The expertise could be due to domain-specific training (e.g., the study of Boolean, propositional logic), or it could be due to general inferential abilities that covary with education (e.g., mental manipulation of abstract symbols). We do not take a position on whether the inference rules are on the order of Piaget's formal logic system, or of a natural logic system (Braine, 1978). The assumption is that these general inferential abilities may allow the subjects to draw inferences from propositions in the absence of experientially based data.

To investigate the hypothesis that general inferential abilities may contribute to performance differences, we sampled subjects from three different education groups. The age distributions of each group were roughly equal, and we controlled statistically for specific training in logic. Also we used only abstract stimuli to control for any thematic knowledge differences between groups. Performance differences between groups would reflect differences in general inferential abilities which covary with level of education.

Method

Subjects. The 300 subjects were employees of AT&T Bell Laboratories ($n=257$) and graduate students at the University of Chicago ($n=43$). Subjects were classified into one of three educational levels based on the highest level of schooling completed. In each group the mean (30) and range (22–55) of ages were about the same. All the high school subjects worked at Bell Labs; 48% had high school diplomas and 52% had 2-year associate degrees. These subjects may not be strictly comparable to previous undergraduate populations because most of them will not obtain 4-year college degrees. In the bachelor's group, 43% of the subjects had nontechnical degrees (social science and business) and 57% had technical degrees (engineering and computer science). In the master's group, 20% of the subjects had nontechnical degrees and 80% had technical degrees; 7% of the subjects had PhDs. Approximately 20% of the subjects in the bachelor's and master's groups were drawn from the University of Chicago. All subjects volunteered to participate.

Design. The experiment employed a 3×4 between-subjects factorial design. The education factor had three levels: high school, bachelor's, and master's. The cue redundancy factor had four levels, similar to those used in Hoch and Tschirgi (1983): (1) *standard* (the original Wason task); (2) *blank* (a blank card representing $\neg q$); (3) *implicit* (an implicit relation between antecedents and consequents); and (4) *explicit* (an explicit antecedent-consequent relation). Subjects were randomly assigned to answer one of the four ver-

sions of the problem, with the constraint that 25 subjects at each education level received each of the four versions.

Materials. The standard version of the reasoning problem was very similar to the original wording of the problem (Wason, 1966):

Assume that each of the boxes below represents a card lying on a table. Each one of the cards has a letter on the front side. Each one of the cards has a number on the back side. Here is a rule that might apply to the cards:

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 front side up with a letter showing, and two of the cards are back side up with a number showing. Your task is to decide which one or ones of the four cards below *must* be turned over in order to find out whether the rule is true or false. You want to minimize the number of cards that you turn over to check, but you must determine whether the rule is true or false for these four cards. Circle the card or cards which must be turned over.

Illustrations of four cards were displayed, showing A, K, 18, and 5.

The other versions of the problem were constructed through minor wording alterations. In the blank problem, subjects were told that each of the cards had a number or a blank on the back side, and the rule read: "If a card has a vowel on its letter side, then it has an even number on its other side." A blank card represented $\neg q$. In the implicit problem, the rule was changed to read: "If a card has a vowel on its letter side, then it has a number greater than 10 on its number side." The implicit problem relied on the ordinal relation between the consequents, highlighting a clear-cut threshold along a familiar continuum to emphasize the differences between 18 and 5 and the inappropriateness of 5 on the back of A (see Hoch & Tschirgi, 1983), for further details). In the explicit problem, the following was inserted after the third sentence of the standard version: "However, a card with a vowel showing may only have an even number on the back side. A card with a consonant showing may have either an odd or an even number on the back side."

Procedure. The reasoning problems were administered to subjects in small groups (1–3) after they had completed one of several unrelated experiments. The experimental materials were assembled in a four-page booklet. The first page introduced the task. The problem was printed at the top of the second page; illustrations of the four cards were displayed (order counterbalanced across subjects) at the bottom of the page where subjects were to indicate their responses. On the third page, subjects were asked to write down their reasons for turning over or not turning over each of the four cards. They were allowed to refer to the previous page to aid them in formulating their reasons, but were explicitly instructed not to change their responses on the second page; however, they were told that, if they wanted to change their answers for any reason, they could do so on the third page beneath their explanations. Subjects were instructed that they were not required to change their answers, but could do so if they felt they had made a mistake. On the fourth page, subjects answered demographic questions, including their highest level of education and whether they had ever studied logic, either "formal logic" and/or "logic design" used in engineering applications.

Results

Two sets of data were analyzed: (1) first- and second-choice-solution probabilities; and (2) subjects' reasons for selecting or not selecting the different cards. Subjects' reasons were classified into inference categories used by Hoch and Tschirgi (1983) and described below. For both sets of analyses, the dependent variables were categori-

Table 1
Solution Frequencies and Percent Correct for First and Second Selection Occasions

Cue Redundancy	Selection Occasion	Education Level								
		High School			Bachelor's			Master's		
		R	W	C	R	W	C	R	W	C
Standard	1st	2	23	.08	1	24	.04	12	13	.48
	2nd	2	23	.08	3	22	.12	13	12	.52
Blank	1st	1	24	.04	5	20	.20	10	15	.40
	2nd	1	24	.04	7	18	.28	15	10	.60
Implicit	1st	2	23	.08	12	13	.48	12	13	.48
	2nd	4	21	.16	14	11	.56	20	5	.80
Explicit	1st	6	19	.24	9	16	.36	18	7	.72
	2nd	6	19	.24	11	14	.44	19	6	.76
Totals	1st	11	89	.11	27	73	.27	52	48	.52
	2nd	13	87	.13	35	65	.35	67	33	.67

Note—R = right selections. W = wrong selections. P(C) = proportion correct.

cal responses and were analyzed by fitting log-linear models estimated by the method of maximum likelihood (Bock, 1975; Bock & Yates, 1973).

Problem solution. Table 1 shows the solution frequencies and proportions for the first- and second-choice occasions according to education and level of cue redundancy. Fifty-two subjects (15 high school, 17 bachelor's, and 20 master's) took the opportunity to change their original answers. None of the subjects who solved the problem on their first attempts changed their answers. As previously mentioned, performance on standard and modified versions of the Wason task using abstract stimuli typically has been quite low, around 10% correct. However, a salient feature of the current data and that of Hoch and Tschirgi (1983) is that performance was relatively high, 38% correct overall on second choices. Moreover, the claim that the problem "is extremely difficult for even highly intelligent subjects" (Griggs, 1983, p. 17) appears overstated: 48% of the master's solved the original (standard) Wason task on their first attempt.

For the first-choice data, there were significant main effects of education [$\chi^2(2) = 44.44$, $p < .001$] and cue redundancy [$\chi^2(3) = 12.05$, $p < .025$]. Subjects with master's degrees performed better than those having bachelor's degrees [$\chi^2(1) = 14.19$, $p < .001$], and the bachelor's subjects performed better than the high school subjects [$\chi^2(1) = 8.94$, $p < .01$]. Planned contrasts, comparing each of the other cue redundancy conditions to the standard version, showed that performance improved significantly with explicit cues [$\chi^2(1) = 11.86$, $p < .001$], and marginally with implicit cues [$\chi^2(1) = 4.76$, $p < .10$]. However, testing of simple main effects (Winer, 1971) indicated that the cue-redundancy factor was significant only within the bachelor's group. The second-choice data were similar to the first-choice data with several exceptions. The chi-square values for all the previously mentioned effects were somewhat larger, especially when comparing the implicit condition to the standard condition [$\chi^2(1) = 14.12$, $p < .001$]. Moreover, simple main effects tests of cue redundancy within the three education levels revealed significant differences for

the bachelor's groups [$\chi^2(1) = 9.36$, $p < .005$] and the master's group [$\chi^2(1) = 5.48$, $p < .05$], with only marginal differences for the high school group [$\chi^2(1) = 4.37$, $p < .10$].

To better understand the differences between the first- and second-choice data, Table 2 summarizes the data for all subjects who did not solve the problem on their first attempts. Three outcomes were possible after these subjects explained their selection decisions for each of the four cards: (1) changing to the correct answer (right); (2) changing to another incorrect answer (wrong); and (3) not changing an incorrect answer (no change). In analyzing these data, the wrong and no-change categories were collapsed into one category (wrong after second choice) and were then compared to the number of correct responses. (The results were the same when the data were analyzed as three separate categories.) Log-linear models were fit to the percent correct figures in Table 2. There were significant main effects of education [$\chi^2(2) = 23.29$, $p < .001$] and cue redundancy [$\chi^2(3) = 11.62$, $p < .025$]. Master's-level subjects were more likely to solve the problem after explaining their selections than bachelor's subjects [$\chi^2(1) = 7.42$, $p < .025$] and bachelor's subjects were more likely than high school subjects [$\chi^2(1) = 5.65$, $p < .05$]. Moreover, the implicit problem led to more correct second choices than did the standard problem [$\chi^2(1) = 9.87$, $p < .005$].

Table 2
Solution Rates on the Second Choice Occasion

Cue Redundancy	Education Level							
	High School		Bachelor's		Master's		Totals	
	C	N	C	N	C	N	C	N
Standard	.00	23	.08	24	.08	13	.08	60
Blank	.00	24	.10	20	.33	15	.12	59
Implicit	.09	23	.18	13	.62	13	.24	49
Explicit	.00	19	.14	16	.14	7	.07	42
Totals	.02	89	.11	73	.31	48	.12	210

Note—Subjects who solved the problem on their first attempt are not included. N refers to the number of subjects who could have changed their wrong first answer. P(C) = proportion correct (no. of correct second answers)/N.

Table 3
Frequencies of Combinations and Individual Cards
for Final Selections

10. Final Selections														
Selection	Education Level													
	High School				Bachelor's				Master's				Total	
	S	B	I	E	S	B	I	E	S	B	I	E		
Combinations														
p, ¬q	2	1	4	6	3	7	14	11	13	15	20	19		115
p	3	12	4	4	5	5	5	3	5	1	1	2		50
p,q*	10	3	6	2	7	6	4	3	3	5	0	0		49
p, ¬p,q, ¬q†	6	1	1	0	4	2	1	0	2	2	2	0		21
other	4	8	10	13	6	5	1	8	2	2	2	4		65
Individual Cards														
p	22	21	18	16	22	22	25	20	24	25	25	23		263
¬p	8	7	9	7	6	5	2	3	3	3	2	0		55
q	20	8	14	7	15	10	5	7	7	8	4	2		106
¬q	6	7	7	12	10	12	15	15	16	19	24	23		166

Note—S = standard; B = blank; I = implicit; E = explicit. *Matching bias. †Biconditional.

Table 3 shows a detailed breakdown of the selection frequencies for combinations and individual cards. The two most common errors were selecting p alone or p and q, the so-called matching bias (Evans & Lynch, 1973). The matching bias, which has been called "the response of last resort" (Hoch & Tschirgi, 1983) and "cognitive short-circuiting" (Yachanin & Tweney, 1982), occurs mainly with the standard problem at lower levels of education. It seems that these subjects do not have much understanding of task demands (Cohen, 1981). As in other studies, using both concrete (Griggs & Cox, 1982) and abstract (Hoch & Tschirgi, 1983) stimuli, the matching bias is reduced when redundant information is added. Selection probabilities of individual cards were also analyzed. More educated subjects made more appropriate selection decisions, correctly not selecting -p [$\chi^2(2) = 16.82$, $p < .001$] and q [$\chi^2(2) = 18.04$, $p < .001$], while correctly selecting -q [$\chi^2(2) = 56.62$, $p < .001$]. Cue redundancy improved performance by reducing q selections [$\chi^2(3) = 23.88$, $p < .001$] and increasing -q selections [$\chi^2(3) = 12.87$, $p < .01$].

The solution data were reanalyzed by including another factor, previous study of logic, in the basic education by cue-redundancy design.¹ As would be expected, subjects with higher levels of education were more likely to indicate that they had studied logic: 16% of high school subjects, 40% of bachelor's subjects, and 59% of master's subjects. That is, exposure to formal logic covaries with educational level. However, the logic/no logic term was not significant [$\chi^2(1) = 3.05$, $p > .15$] when included in the cue-redundancy and education model. Although the cue-redundancy and education model fit the data very well [$\chi^2(6) = 4.62$, $p = .59$], the logic/no logic and cue-redundancy model yielded a poor fit [$\chi^2(19) = 67.85$, $p = .001$]. When level of education was controlled for, there were no significant differences between the logic and no logic groups. There are several possibilities why the logic factor was not as good a predictor of performance as was education level. First, reasoning ability may depend more on general intelligence (as reflected in the ability to

progress in the educational system) than on specific training in logic. However, the result could also be the result of an inadequate measure of logic training (e.g., not taking into account the extent or level of exposure).

In summary, education level had a large impact on reasoning performance. Moreover, redundancy in the form of implicit or explicit cues increased the probability of solution.² This improvement was greatest for the bachelor's subjects, which may partially reflect floor and ceiling measurement effects in the high school and master's groups, respectively. Education level also increased the probability that subjects would detect an error in their initial choice and solve the problem on their second try. Subjects answering the implicit problems were also more likely to detect and then correct faulty reasoning. Although explicit problems were solved at about the same rate as implicit problems, very few subjects changed their initial answers, indicating that the explicit cues were either detected on the first attempt or not at all.

Reasons for selections. Subjects' reasons were categorized into the inference categories developed by Hoch and Tschirgi (1983). An inference is defined as any statement about the antecedents, consequents, or the rule that was not explicitly stated in the problem but was inferred from the available information. Although retrospective protocols may contain elements of rationalization (Evans & Wason, 1976) and may not always reflect the true psychological processes underlying performance, they can supplement the more objective performance data. As in Hoch and Tschirgi (1983), we were interested in understanding the basis of the cue-redundancy effect. All of the reported log-linear analyses are based on the data excluding the explicit conditions, because in these problems subjects were told that -p implies both q and -q, truth values that they would have to infer in the other conditions. The purpose of these analyses was to determine which conditions would lead subjects to make extra-logical inferences. From our previous study, we expected that the implicit conditions would lead to a pattern of extra-logical inferences similar to that found in the reasons data of the ex-

PLICIT subjects. Inferences about p were not analyzed; because 88% of the subjects correctly selected the p card, these data did not discriminate between solvers and non-solvers.

-p inferences. Most of the reasons accompanying a correct decision to not check $-p$ fell into two inference categories. Explanations were coded as " $-p$ -irrelevant" if subjects explicitly stated that $-p$ was not a part of the rule and therefore not relevant. Explanations were coded as " $-p$ implies both q and $-q$ " if subjects stated that either q or $-q$ were appropriate with $-p$. Although the $-p$ -irrelevant and $-p$ -implies-both inferences appear similar, only the second inference is congruent with the truth table; the first inference may be symptomatic of a lack of generating all antecedent-consequent pairs. An "other" category included the few remaining explanations of subjects who correctly did not select $-p$ and all the explanations of those incorrectly selecting $-p$.

Table 4 shows that inference production was influenced by both education [$\chi^2(4) = 33.40, p < .001$] and cue redundancy [$\chi^2(4) = 15.86, p < .005$]. The number of $-p$ -implies-both inferences increased with higher levels of education; $-p$ -irrelevant inferences were equally likely across education levels. Subjects answering the implicit problem were more likely to infer that $-p$ implies both

and less likely to conclude that $-p$ is irrelevant than were subjects in either the standard or blank conditions. The pattern of inferences generated by implicit subjects was very similar to the reasons data of explicit subjects, even though explicit subjects were specifically provided with the information that $-p$ implies both q and $-q$.

Table 5 shows the solution frequencies for the three different $-p$ inferences.³ Although 70% of the subjects making $-p$ -implies-both inferences solved the problem, only 21% of those concluding $-p$ is irrelevant and 17% of those giving other explanations were correct [$\chi^2(4) = 52.34, p < .001$]. The $-p$ -implies-both inference is the logically correct inference, but as Wason and Johnson-Laird (1972) have stated, the $-p$ -irrelevant inference, by itself, in no way precludes solution of the problem. It is only necessary that subjects recognize (p, q) as a legal pair and $(p, -q)$ as an illegal pair. However, in this study as in Hoch and Tschirgi (1983), the $-p$ -irrelevant inference may be symptomatic of misperception of other antecedent-consequent pairings. This conclusion is supported by examination of the $-q$ inferences that accompanied each of the $-p$ inferences. Sixty percent (49/80) of the subjects who stated $-p$ -irrelevant also stated $-q$ -irrelevant, a conclusion that guarantees an incorrect answer; subjects inferring $-p$ -irrelevant made the correct decision about $-p$,

Table 4
Frequencies of Extra Logical Inferences for $-p$, q , and $-q$ by
Level of Education and Cue Redundancy

Cue Redundancy	-p Inferences			q Inferences			-q Inferences			
	-p Implies	Both	Irrelevant	Other	q Implies	Both	Other	p, -q Violates	-q Irrelevant	Other
High School										
Standard	1		9	15	1		24	5	9	11
Blank	5		11	9	3		22	4	10	11
Implicit	5		5	15	3		22	5	6	14
Explicit	9		4	12	10		15	9	3	13
Bachelor's										
Standard	6		8	11	8		17	10	6	9
Blank	8		10	7	10		15	12	10	3
Implicit	15		6	7	16		9	16	8	1
Explicit	18		4	3	16		9	16	3	6
Master's										
Standard	10		11	4	17		8	16	6	3
Blank	13		7	5	15		10	19	4	2
Implicit	17		4	4	21		4	23	1	1
Explicit	20		5	0	20		5	23	1	1

Table 5
Number of Subjects Who Were Correct and Incorrect by
Level of Education and Cue Redundancy

Selection	$\neg p$ Inferences			q Inferences			$\neg q$ Inferences			
	$\neg p$			q			$p, \neg q$	$\neg q$		
	Implies	Both	Irrelevant	Other	Implies	Both	Other	Violates	Irrelevant	Other
All Problems										
Correct	81		18	16	98		17	110	2	3
Incorrect	35		75	75	39		146	48	65	72
Excluding Explicit Problems										
Correct	49		17	13	66		13	74	2	3
Incorrect	21		63	62	25		121	48	58	40

but many of those decisions (at least 60%) appear to have been based on the wrong reason. By contrast, only 3% (2/70) of those inferring $\neg p$ -implies-both incorrectly concluded $\neg q$ -irrelevant. Subjects who inferred that a card was irrelevant (whether $\neg p$ or $\neg q$) may have dismissed these cards without actually considering symbols that might have been on the other side, (i.e., the reversibility problem, Wason & Johnson-Laird, 1972).

q inferences. There was one predominant inference accompanying the correct decision to not select q . Explanations were coded as " q implies both p and $\neg p$ " if subjects stated that either p or $\neg p$ were appropriate with q . An "other" category consisted of all other explanations for either correct or incorrect selections. Inference generation differed across education [$\chi^2(2) = 67.56, p < .001$] and cue redundancy [$\chi^2(2) = 8.63, p < .025$]. Increasing levels of education led to greater numbers of q -implies-both inferences. Also subjects were more likely to infer that q -implies-both when answering implicit problems. Again, the inference patterns were very similar for the implicit and explicit conditions, except for high school subjects whose reasons were not influenced by the implicit cue.

Although 73% of those subjects inferring q -implies-both solved the problem, only 10% of those generating other reasons solved the problem. Even though there are other reasons that subjects might decide to not select q , the q -implies-both inference indicates understanding of two of the four antecedent-consequent pairings. Table 6 shows the solution frequencies conditional on $\neg p$ and q inference patterns. Those subjects making both extra-logical inferences solved the problem 88% of the time; 42% of those subjects generating only one of the inferences were correct; and 7% of those making neither inference were correct [$\chi^2(2) = 115.31, p < .001$].

$\neg q$ inferences. There was one predominant explanation accompanying the correct decision to select $\neg q$. Most subjects said that the rule would be violated if p appeared on the back of $\neg q$, or that only $\neg p$ could appear on the back of $\neg q$. We combined these explanations into one inference category, " p with $\neg q$ violates the rule." Subjects also stated that $\neg q$ was irrelevant, leading to an incorrect selection decision. An "other" category consisted of all the other explanations for correct and incorrect selections. There was a strong effect of education [$\chi^2(4) = 63.47, p < .001$], but little effect of cue redundancy [$\chi^2(6) = 7.79, p < .10$] on the types of $\neg q$ inferences.

Table 6
Number of Subjects Providing Extralogical Explanations
Who Solved and Did Not Solve the Task

	All Problems			Excluding Explicit Problems		
	Both	Either	Neither	Both	Either	Neither
Correct	83	18	12	51	19	9
Incorrect	16	36	135	7	26	113

Note—Both = $\neg p$ implies both and q implies both. Either = $\neg p$ implies both or q implies both. Neither = neither $\neg p$ implies both nor q implies both.

Higher educated subjects more often stated that p -violates and did not falsely conclude $\neg q$ -irrelevant.

Overall, subjects' explanations of their selection decisions were similar to those in Hoch and Tschirgi (1983); however, salient differences arose in the blank condition. In the previous study, we found that blank problems led to more p -violates inferences for the $\neg q$ card. From these data, we suggested that the blank facilitated reasoning performance by cuing the counterexample ($p, \neg q$). The current data do not show this facilitating effect of blank cards, however.

REPLICATIONS

The four-card selection task (Wason, 1966) was an early demonstration of the difficulty humans have in detecting and using the logical structure of material implication. The previous study provides a rigorous replication of the positive effect of redundant cues to logical structure even when the true-false version of the task and abstract stimuli are used. We had hypothesized that large performance differences would emerge for the different educational levels; however, the 48% solution rate on the standard task for master's subjects was a surprise. Moreover, several other researchers have found poor performance by college-level faculty on standard versions (Griggs & Ransdell, 1985; Kerns, Mirels, & Hinshaw, 1983; cf. Tweney & Yachnin, 1985). Three replications were run to ensure that the performance of the master's subjects was a reliable finding, and to isolate its basis by obtaining more detailed information about background logic knowledge and formal training.

Method

Two versions of the standard task were used to assess whether the improvement might have been due to minor wording differences between our stimulus and other researchers' stimuli. Two groups of master's-level subjects from Bell Labs ($n=25$ each) were given either our standard version or a verbatim reproduction of the "testing/abstract" stimulus from Griggs and Ransdell (1985). Most (94%) of these subjects had master's degrees in either engineering or computer science/math. A third group of graduate students at the University of Chicago ($n=25$) were given our standard version; these subjects had master's degrees in economics/business (72%) or in life sciences (28%). The procedure from the previous study was used, with the addition of a set of questions to gather more detailed information about prior knowledge. Subjects were asked whether they had ever seen or solved a similar problem and where. Also, they rated their knowledge of five topics deemed relevant to deductive reasoning: Boolean logic, the scientific method, propositional logic, circuit design, and formal logic. The scale ranged from 1 (not knowledgeable, "I could barely define the term") to 5 (very knowledgeable, "I could teach a course on the subject"). In retrospect, we probably should have measured prior knowledge before completing the task, because these self-report data may in part represent confidence ratings about task performance.

Results

The results for first choices were as follows. Our standard version was solved by 40% (10/25) of the first Bell Labs sample; the Griggs and Ransdell (1985) version was solved by 40% (10/25) of the other Bell Labs sample. In the University of Chicago sample, 44% (11/25) of the sub-

jects solved our standard version of the task. Solution rates after subjects' explanations increased in each sample: 64% (16/25), 52% (13/25), and 48% (12/25), respectively. These data provide strong corroboration of good performance by master's-level subjects, at least for the two populations studied here.

In an effort to isolate the basis for this good performance, a variety of analyses were conducted to test for systematic individual differences between solvers and nonsolvers. First, were these subjects naive to the Wason task? Twenty-five percent of the subjects (19/75) said that they had seen or solved a "problem of this form" in "puzzle books," "standardized tests," "paradoxical jokes," and "introductory logic courses." This apparent lack of task naivete might have been troublesome, except for the fact that these subjects were no more likely to solve the problem than the rest of the sample [58% (11/19) of the "nonnaive" subjects vs. 54% (30/56) of the "naive" subjects]. Ratings of prior knowledge of the five topics relevant to deductive reasoning also were analyzed. The ratings data constituted the dependent variables in a MANOVA; solvers/nonsolvers and replication group were the between-subjects independent variables. The two Bell Labs groups gave much higher self-ratings than did the University of Chicago subjects [mean = 2.74 and mean = 2.91 vs. mean = 1.72, $F(10,130) = 5.52$, $p < .001$]. This difference is interesting because there were no performance differences between these groups. The solver/nonsolver variable was analyzed using both first- and second-choice data. In both cases solvers and nonsolvers did not differ significantly in terms of prior knowledge [first choices $F(5,68) < 1$, and second choices $F(5,68) = 1.79$, $p = .13$]. Discriminant analysis using solvers/nonsolvers as the grouping variable also identified no systematic differences.

Because of the disparate performance of these master's-level subjects compared to results for previously studied populations, a few additional comments are in order. First, our subjects were a highly self-selected and motivated group with (we suspect) high intelligence, probably similar to the "scientist" subjects in previous studies. Second, these master's subjects undoubtedly have well developed analytical skills. Specifically, they may be more accustomed to thinking and reasoning at an abstract level. Clearly, they have progressed through an educational regimen which required expression and comprehension of abstract (mathematical) symbols; we speculate that master's-level subjects in other fields (e.g., English or education) may not be as accomplished or practiced at abstract symbol manipulation, and therefore may not display the same level of competence. Finally, we take comfort in the fact that, after 20 years, a group of subjects has been found that displays some competence in reasoning about Wason's selection task.

GENERAL DISCUSSION

Brunswik's model of perception (1952) as a process of the integration of multiple probabilistic cues provides a

general framework for understanding previous selection task studies in terms of cue redundancy. Memory-cuing problems such as Griggs and Cox's (1982) drinking-age task provide a redundant way to recognize the truth values of all the combinations of beverages and ages; moreover, subjects' personal or vicarious experiences with the regulation may make the counterexample (underage beer drinkers) highly salient. Other problems relying on long-term memory-cuing (e.g., Johnson-Laird, Legrenzi, & Sonino-Legrenzi, 1972) can be understood in a like manner. Reasoning by analogy can be interpreted as a case in which redundancy transfers from a previously presented problem. However, it is important to recognize the differences between these problems and those used here and in Hoch and Tschirgi (1983). Problems that cue long-term memory conceivably could have been solved without any logical knowledge. Such is not the case in our own studies. Subjects had to be able to effectively coordinate both logical (implication rule) and extra-logical (redundant-relational information) cues to solve the problem. As this study demonstrates, however, redundant cues are not sufficient by themselves to allow subjects to solve the abstract problem. The implicit and explicit relational cues provide useful information only if subjects have some knowledge about the logical structure of conditionals.

The large differences in performance across education level highlight the importance of a subject's access to some form of inference rules or general ability to reason logically. Providing redundant cues to high school subjects did not improve performance to any great degree. For the most part, they appeared unable to access the necessary logical structure to benefit from the redundant cues. The most dramatic improvement in reasoning occurred for the bachelor's group. When these subjects had to depend solely on the logic cue (the standard version), they could access only a partial representation of the problem structure. In contrast, redundant cues helped these subjects to fill in the gaps or inaccuracies in their psychological truth tables, thereby demonstrating competency with material implication (see Hammond, Hamm, Grassia, & Pearson, 1984, for a discussion of the importance of the congruence between task structure and the task strategies adopted by subjects). By comparison, many of the master's-level subjects had more complete representations to start with; for over 50% of these subjects in the standard condition, the implication rule provided sufficient information to solve the problem. This is not to say that master's subjects did not use the redundant cues when they were available; the explanation data indicate that they did. Moreover, in other studies using thematic content, subjects may have relied even more heavily on the structural relations implicated by the redundant cues rather than whatever logical knowledge they did possess (D'Andrade, 1982; Hoch & Tschirgi, 1983).

The second-choice data provide additional support for the importance of cue redundancy. Subjects with more education had more difficulty "rationalizing" inconsistent selection decisions, especially on implicit problems. Eighty-eight percent of the subjects who did not solve the

problem on their first attempts did not catch their mistakes while explaining their choices. This was not the case for the master's subjects in the implicit condition, where vicariously functioning cues helped 8/13 (62%) recognize initially faulty reasoning.

Recently, there have been a number of proposals that human reasoning is example-based rather than guided by formal rules of inference (Cohen, 1981; D'Andrade, 1982; Griggs, 1983; Mandler, 1980; Rumelhart, 1980). These proposals and the research that supports them (e.g., Shaklee, 1979; Trabasso, 1977; Tschirgi, 1980) are partly a reaction against the strong Piagetian position (Inhelder & Piaget, 1958) concerning the development of formal operations in childhood and adolescence. At one extreme, Piaget (1972) argued that the ability to reason verbally (as opposed to the manipulation of concrete objects or their representations) requires formal operations, the ability to reason hypothetically, independent of the intrinsic truth of the premises. At the other extreme are proposals that all reasoning knowledge is context-dependent, "tied to particular schemata related to particular bodies of knowledge" (Rumelhart, 1980, p. 55), or embedded in "task-specific procedures rather than in general rules of inference" (Mandler, 1980, p. 32). The results of the present experiment and Hoch and Tschirgi (1983) suggest that an understanding of human reasoning requires recognition that subjects may be able to draw on both formal and example-based representations of logical structure. Future research on human reasoning must recognize the interrelationship between logical and extralogical cues in order to understand how individual subjects can capitalize on their own knowledge to reason in an effective manner.

REFERENCES

- BOCK, R. D. (1975). Multivariate analysis of qualitative data. In R. D. Bock (Ed.), *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.
- BOCK, R. D., & YATES, G. (1973). *MULTIQUAL: Log-linear analysis of nominal or ordinal qualitative data by method of maximum likelihood*. Chicago: National Educational Resources.
- BRAINE, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, **85**, 1-21.
- BRUNSWIK, E. (1952). The conceptual framework of psychology. In *International encyclopedia of unified science* (Vol. 1, No. 10). Chicago: University of Chicago Press.
- COX, J. R., & GRIGGS, R. A. (1982). The effects of experience on performance in Wason's selection task. *Memory & Cognition*, **10**, 496-502.
- COHEN, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral & Brain Science*, **4**, 317-370.
- DAWES, R. M. (1975). The mind, the model, and the task. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory* (Vol. 1). Hillsdale, NJ: Erlbaum.
- D'ANDRADE, R. (1982, April). *Reason versus logic*. Paper presented at the Symposium on the Ecology of Cognition: Biological, Cultural, and Historical Perspectives, Greensboro, NC.
- EINHORN, H. J., & HOGARTH, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, **85**, 395-416.
- EVANS, J. ST. B. T. (1982). *The psychology of deductive reasoning*. London: Routledge & Kegan Paul.
- 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., & WASON, P. C. (1976). Rationalization in a reasoning task. *British Journal of Psychology*, **67**, 479-486.
- GRIGGS, R. A. (1983). The role of problem content in the selection task and the THOG problem. In J. St. B. T. Evans (Ed.), *Thinking and reasoning: Psychological approaches*. London: Routledge & Kegan Paul.
- GRIGGS, R. A. (1984). Memory cueing and instructional effects on Wason's selection task. *Current Psychological Research & Reviews*, **3**, 3-10.
- GRIGGS, R. A., & COX, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, **73**, 407-420.
- GRIGGS, R. A., & RANDELL, S. E. (1985). *Scientists and the selection task*. Unpublished manuscript, Department of Psychology, University of Florida, Gainesville.
- HAMMOND, K. R. (1966). Probabilistic functionalism: Egon Brunswik's integration of the history, theory, and method of psychology. In K. R. Hammond (Ed.), *The psychology of Egon Brunswik*. New York: Holt, Rinehart & Winston.
- HAMMOND, K. R., HAMM, R. M., GRASSIA, J., & PEARSON, T. (1984). *The relative efficacy of intuitive and analytical cognition: A second direct comparison*. (Report No. 252). Boulder: University of Colorado, Center for Research on Judgment and Policy.
- HOCH, S. J., & TSCHIRGI, J. E. (1983). Cue redundancy and extra logical inferences in a deductive reasoning task. *Memory & Cognition*, **11**, 200-209.
- INHELDER, B., & PIAGET, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.
- JOHNSON-LAIRD, P. N., LEGRENZI, P., & SONINO-LEGRENZI, M. (1972). Reasoning and sense of reality. *British Journal of Psychology*, **63**, 395-400.
- JOHNSON-LAIRD, P. N., & TAGART, J. (1969). How implication is understood. *American Journal of Psychology*, **82**, 367-373.
- KERNS, L. H., MIRELS, H. L., & HINSHAW, V. G. (1983). Scientists' understanding of propositional logic: An experimental investigation. *Social Studies of Science*, **13**, 131-146.
- MANDLER, J. M. (1980). *Structural invariants in development*. (Tech. Rep. No. 76). San Diego: University of California, Center for Human Information Processing.
- PIAGET, J. (1972). Intellectual development from adolescence to adulthood. *Human Development*, **15**, 1-12.
- RUMELHART, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Erlbaum.
- SHAKLEE, H. (1979). Bounded rationality and cognitive development: Upper limits on growth? *Cognitive Psychology*, **11**, 327-345.
- TRABASSO, T. (1977). The role of memory as a system in making transitive inferences. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 333-366). Hillsdale, NJ: Erlbaum.
- TSCHIRGI, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, **51**, 1-10.
- TWENEY, R. D., & YACHANIN, S. A. (in press). Can scientists rationally assess conditional inferences? *Social Studies of Science*.
- VAN DUYN, P. C. (1976) Necessity and contingency in reasoning. *Acta Psychologica*, **40**, 85-101.
- WASON, P. C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology*. London: Penguin.
- WASON, P. C. (1983). Realism and rationality in the selection task. In J. St. B. T. Evans (Ed.), *Thinking and reasoning: Psychological approaches*. London: Routledge & Kegan Paul.
- WASON, P. C., & JOHNSON-LAIRD, P. N. (1972). *Psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.
- WINER, B. J. (1971). *Statistical principles in experimental design* (2nd. ed). New York: McGraw-Hill.
- YACHANIN, S. A., & TWENEY, R. D. (1982). The effect of thematic content on cognitive strategies in the four-card selection task. *Bulletin of the Psychonomic Society*, **19**, 87-90.

NOTES

1. The logic/no logic groups were based on whether subjects indicated previous study of "formal logic." The analyses turned out the same when the logic/no logic partition was based on the study of "logic design" or on a conjunction or a disjunction of formal logic and logic design.

2. We reasoned that subjects performed better on the implicit problem because the "greater than" relation established a distinct subordinate relationship between the consequents (analogous to 22- vs. 16-year olds in the drinking-age problem). However, another possibility is that performance was improved by eliminating the opportunity for a matching bias response—the q (10) mentioned in the rule was not the same as the q (18) shown on the individual cards. To test this idea, we constructed

a variant of the implicit problem in which matching bias could operate—the rule was changed to read "greater than or equal to 10" and the q card showed a "10." Sixty percent (15/25) solved the "greater than" version, and 56% (14/25) solved the "greater than or equal to" version, suggesting that eliminating the opportunity for a matching bias response is not what leads to improved performance in the implicit condition.

3. The reported analyses are based on the second-choice data, because the reasons of subjects who changed their answers typically reflected those second choices.

(Manuscript received December 11, 1984;
revision accepted for publication June 10, 1985.)