Cue redundancy and extra logical inferences in a deductive reasoning task

STEPHEN J. HOCH

Graduate School of Management, Northwestern University, Evanston, Illinois 60201

and

JUDITH E. TSCHIRGI

Bell Telephone Laboratories, Inc., Naperville, Illinois 60566

This study investigated the influence of extraexperimental knowledge on deductive reasoning. Two groups of subjects solved modified abstract or concrete versions of Wason's (1966) four-card selection problem in which the "if-then" implication rule provided the sole cue to underlying logical structure. Six other groups solved abstract or concrete comparison problems in which additional information was supplied by varying the relationship between antecedents and consequents in a manner consistent with logical structure. Selection responses and subjects' explanations were analyzed using log-linear models of data arrayed in multiway contingency tables. Results showed improved performance for concrete over abstract problems and for both abstract and concrete problems that included relational information. Subjects capitalized upon the redundancy between the implication rule and the relational cues to reason effectively with either abstract or concrete stimuli.

In 1966, Wason first reported subjects' difficulty with a deductive reasoning task that has come to be known as the "Wason four-card selection task." In the task's original form, subjects were given an implication rule of the form "if p then q " along with a set of four cards. On one side of each card was an antecedent, either p or $\sim p$ (not p), and on the other side was a consequent, either q or $\sim q$ (not q). The antecedents were letters, E or K, and the consequents were numbers, 4 or 7. The rule was "if a card has a vowel on one side, then it has an even number on the other side." The task was to decide which of four cards (showing a p , $\sim p$, q , or $\sim q$) must be turned over to determine the truth or falsity of the rule. The necessary and sufficient solution was p and $\sim q$. This was because, of the four possible antecedent-consequent pairs, only the $(p, \sim q)$ pair violated the rule. Subjects, however, typically selected p and q , termed a "verification bias," or, less often, p alone. Evans and Lynch (1973) showed that this apparent bias to search for confirming instances in the four-card selection task was due to a tendency for subjects to select those items mentioned in the rule, a matching bias.

While selection task performance remained very low over several replications (typically less than 10% of the subjects were correct), Johnson-Laird, Legrenzi, and Sonino-Legrenzi (1972) and Wason and Shapiro (1971) found improved performance when concrete, thematic materials were substituted for the letters and numbers. Wason and Shapiro used a task with cities as the antecedents and modes of travel as the consequents; Johnson-Laird et al. had subjects imagine they were postal

workers examining sealed and unsealed letters with stamps of different denominations. Since these two studies, however, the support for improved performance with thematic materials has been equivocal (see Griggs & Cox, 1982, for a comprehensive summary).

Griggs and Cox (1982) replicated the Johnson-Laird et al. (1972) postal study and found no thematic effect. However, they did uncover an old British postal regulation stating that sealed envelopes required higher priced stamps than unsealed envelopes. A lower priced stamp ($\sim q$) on a sealed envelope (p) would be a clear violation of the regulation. Griggs and Cox argued that the impressive performance in the earlier study occurred because subjects were relying upon their memory for the postal regulation as a cue to recognize a falsifying instance. Griggs and Cox's subjects were Florida undergraduates who could not use British postal regulations to recognize a false instance. Griggs and Cox then constructed a problem in which the implication rule concerned the legal drinking age in Florida, part of their subjects' past experience. The rule was "If a person is drinking beer, then the person must be over 19 years of age." The four card items were "beer," "Coke," "22 years," and "16 years." A 16-year-old drinking beer was obviously a violation that subjects would recognize quickly from first- or second-hand experience. Indeed, subjects given these thematic materials performed much better than those given abstract problems. Griggs and Cox concluded that the thematic materials effect was due to long-term memory cuing from "mundane experience," in which problem scenarios matching past experi-

ence or knowledge might allow subjects to construct counterexamples, the $(p, \sim q)$ case. Subjects could use this counterexample as a cue to solve the four-card problem without necessarily recognizing all the logical implications of the task.

The current research addresses two related questions. First, why do subjects have so much trouble with abstract versions of the selection task? The material implication of the conditional "if p then q " is sufficient to infer the logical structure underlying the selection task as formally represented by the truth table [i.e., (p, q) , $(\sim p, q)$, and $(\sim p, \sim q)$ are true pairs and $(p, \sim q)$ is false]. When Johnson-Laird and Tagart (1969) presented subjects with the four pairs, subjects stated that (p, q) was an allowable pair and $(p, \sim q)$ was an illegal pair, but they indicated that both $(\sim p, q)$ and $(\sim p, \sim q)$ were "irrelevant." Despite this "defective truth table," Wason and Johnson-Laird (1972) contended that subjects could have solved the selection task with only their knowledge of the truth value of (p, q) and $(p, \sim q)$. The fact that subjects still failed to select the $\sim q$ card has been ascribed to a lack of understanding of the reversibility of the $(p, \sim q)$ relation (Johnson-Laird & Wason, 1970), that is, that p on the other side of $\sim q$ is just as illegal as $\sim q$ on the other side of p . Unlike the Johnson-Laird and Tagart task, the selection task would seem to require that people first generate all possible pairwise card combinations and then identify the truth value of each pair. We believe that the reversibility problem occurs because subjects have difficulty in spontaneously generating the two potential antecedent-consequent pairs for each of the four different items (p , $\sim p$, q , $\sim q$) that might be face up. Although Johnson-Laird and Tagart showed that people could identify the truth value of enough pairs to solve the problem, we do not know how easily they could have generated those four pairs plus the reverse of those pairs on their own.

The second question is what is the basis of the facilitation due to thematic materials? Most people cannot use the implication rule alone to deduce the information in a truth table. The rule most often cues the (p, q) pair and its true value (i.e., a concern with confirming that q is indeed on the back side of p). We hypothesize that thematic content improves performance when it provides alternate cues for the generation of other pairs and the identification of their truth values. Thematic material can provide redundant cues to logical structure (Brunswik, 1952; Hammond, 1966). Subjects no longer have to rely solely upon the implication rule to solve the problem but may use "vicariously functioning" cues to infer allowable and nonallowable relations between antecedent-consequent pairs. For example, a familiar counterexample such as a 16-year-old illegally drinking beer provides a redundant way to generate the antecedent-consequent pair $(p, \sim q)$ and to recognize it as a violation of the rule (i.e., the drinking age in Florida). Counterexamples highlight illegal rela-

tionships between antecedent-consequent pairs. Griggs and Cox (1982) and Manktelow and Evans (1979) also have argued that solution of thematic problems need not be by strict logical implication. They imply, however, that subjects invoke domain-specific, personally experienced knowledge to construct counterexamples to the rule. We believe that people can construct counterexamples from more general antecedent-consequent relations not specific to a thematic domain. We expect people can generate counterexamples in abstract problems as well, and thus identify the false antecedent-consequent pair.

We further hypothesize that redundant cues that help subjects generate and identify true pairs [other than (p, q)] can also improve performance. Recall that the most common mistake in the abstract problems is to select the q card, indicating that subjects are concerned about the value on the other side. We know from the Johnson-Laird and Tagart (1969) task that subjects do not identify the $(\sim p, \sim q)$ and $(\sim p, q)$ cases as true but, rather, as irrelevant to the task. If subjects recognized that both the (p, q) and $(\sim p, q)$ pairs were true, there would be no reason for them to turn over the q card, since they would realize that q was appropriate with both values of p . We suggest that propositional content can cue subjects to the truth of the $(\sim p, \sim q)$ and $(\sim p, q)$ pairs. These inferences should help subjects to avoid the most common selection error, turning over q , and therefore increase the probability of a correct solution.

A pilot experiment provided preliminary support for our prediction that cues about both allowable and nonallowable pairwise relations would aid performance. We constructed thematic and abstract problems similar to those used by D'Andrade (Note 1). D'Andrade had constructed two selection problems. In one problem, he asked subjects to imagine that they were workers in the Pica Custom Label Company, inspecting labels that had letters, E and K, printed on one side and numbers, 3 and 8, on the other. The rule was "If a vowel, then an odd number." In the other problem, subjects were told to imagine that they worked at Sears, checking sales receipts to make sure that any sale over \$30 had been approved by a department manager. The subjects then had to decide whether to check receipts showing \$75, \$25, a manager's approval signature, or no approval signature. Seventy percent of the subjects solved the Sears problem, correctly checking \$75 and the unsigned sales receipt. Only 13% solved the Pica Label problem, correctly checking E and 8.

We thought that the improved performance on the Sears problem might be due to the presence of a blank sales slip (the $\sim q$ item was represented by a blank card with no signature). It is possible that subjects were relying upon extraexperimental knowledge about situations in which something is absent or missing; that is, if nothing is there, then something must be wrong. This knowledge about situations in which something is miss-

ing is not domain specific; it is a common cue that occurs across many different contexts. Our pilot experiment attempted to disentangle the effect of thematic materials from the "absence" cue. Our hypothesis was that the absence cue should improve performance with both abstract and concrete themes.

In our abstract-standard problem, widget inspectors checked Model A and Model B widgets for series numbers 1 or 2. The rule was "If a widget is a Model A, then there must be a 2 on the back." The four cards showed A, B, 2, or 1. The abstract-blank problem was changed so that it was appropriate to have no series number (a blank) in some cases. The rule was "If a widget is a Model A, there must be a series number on the back." The four cards showed A, B, 2, or blank. The concrete problems involved cashiers checking vouchers for sales personnel at Company X. In concrete-standard problems, cashiers checked for regional and national sales managers' signatures on vouchers. The rule was "If a voucher is for more than \$500, there must be a national manager's signature on the back." The four cards showed "\$1,000," "\$300," "national," or "regional." The concrete-blank problem required that only some vouchers have a manager's signature. The rule was "If a voucher is for more than \$500, there must be a Sales Management signature on the back." The four cards showed "\$1,000," "\$300," "national," or "_____." Both abstract and concrete problems required the subjects to make sure the rule was followed.

We found that performance in both concrete conditions surpassed performance in the abstract conditions, 73% and 25%, respectively. However, subjects in the abstract-blank condition (39%) performed significantly better than subjects in the abstract-standard condition (11%). Interestingly, we found a nonsignificant difference between the two concrete conditions, 84% and 66%, respectively. One possible explanation for this nonsignificant effect was that the voucher-standard problem included uncontrolled extraexperimental cues that facilitated solution. First, the antecedents, \$1,000 and \$300, were related in a salient and easily understood manner: They were ordered dollar amounts that differed from a stated referent, \$500. Second, the voucher consequents, national and regional manager, were also easily ordered in that a national position typically dominates regional in any organization. Moreover, the relationship between the antecedents and the relationship between the consequents covaried positively: \$1,000 was associated with national and \$300 with regional. Given this problem content, it seems likely that even though the information was not explicitly provided, subjects would infer that a national manager, as a superordinate boss, was authorized to sign both the \$1,000 and \$300 vouchers, whereas a regional (subordinate) manager could only sign the \$300 voucher. Making these inferences is equivalent to generating and identifying all allowable and nonallowable antecedent-consequent pairs, resulting in a mental representation compatible with the complete truth table.

If our analysis of problem content is correct, we should find a significant effect of the blank in a concrete problem if we eliminate other interrelations between the cards. A blank card, as a salient counterexample, should help subjects overcome difficulty in generating both the $(p, \sim q)$ and $(\sim q, p)$ pairs (i.e., reversibility). Providing interitem relational cues about $\sim p$ and q should help subjects to realize that every antecedent-consequent pair must be true or false, rather than irrelevant. The addition of such redundant cues should allow subjects to overcome the "defective" truth table (Wason & Johnson-Laird, 1972) that appears to underlie adults' reasoning about a conditional.

To provide a baseline comparison between the abstract and concrete materials, we needed one abstract and one concrete problem in which the only cue to solution resided in the stated rule. We referred to this problem type as "no relation." We used a modified version of the Wason (1966) task for the basic abstract/no-relation problem. The task involved determining whether four instances obeyed a procedural rule, instead of determining the truth or falsity of a rule, as in the original problem.¹ The analogous concrete problem was a variant of the quality-control clerk scenario. Instead of using abstract labels on widgets, we used more realistic stimuli: model numbers and instruction panels on pocket calculators. The situation satisfied the criteria of being realistic (or thematic) but not familiar, unlike the postal problems of Johnson-Laird et al. (1972) and the drinking-age problems of Griggs and Cox (1982). The concrete/no-relation problem was quite similar to van Duyne's (1976) concept of an arbitrary antecedent-consequent relation in the postal scenario.

The first redundant cue was the absence cue. This cue seemed to aid performance because it increased the salience of the $(p, \sim q)$ pair. The p item has always been highly salient for subjects, and most subjects decide to turn over the p item to make sure that q is on the other side. However, most subjects do not realize the reversibility of the $(p, \sim q)$ relationship, not realizing that p on the back of $\sim q$ also violates the rule. When $\sim q$ is represented by "an object with something missing," a red flag is raised, signaling the need for further consideration of the other side of $\sim q$.

The second redundant cue, the implicit cue, is more understandable if we refer to the discussion of the travel voucher problem in the pilot experiment. We assumed that subjects inferred that the regional manager's ($\sim q$) domain of authority was a subset of the national manager's (q) authority. A national manager could sign for both \$300 ($\sim p$) and \$1,000 (p) vouchers, eliminating the need to turn over the national manager (q) card. The criterion we adopted when developing the implicit stimuli was that subjects should infer either that $\sim p$ is allowable with q and $\sim q$ or that q is allowable with p and $\sim p$. Either of these inferences should help subjects to generate and identify the truth of the pairs $(\sim p, q)$ and $(\sim p, \sim q)$.

The final redundant cue involved explicitly providing

subjects with the analogous information that was only implied in the implicit condition. Explicit relation problems included the following information beyond the basic logical format: The p items may only have q on the back, whereas $\sim p$ items may have either a q or a $\sim q$ on the back.

METHOD

Subjects

Two hundred students at the Northwestern Graduate School of Management participated in the study. All subjects were in their 1st or 2nd year of a 2-year MBA program. Nine subjects were replaced because they were not fluent in English.

Design

The experiment was a 2 by 4 factorial design. The theme factor had two levels: abstract and concrete. The relation factor had four levels: no relation, blank, implicit, and explicit. The experiment was run completely between subjects, with each subject randomly assigned to solve one of the eight problems. Twenty-five subjects solved each type of problem.

Procedure

The reasoning problems were administered to groups of 20-30 subjects during regularly scheduled classes. Each problem was self-explanatory and was typed on one side of an 8 x 11 in. sheet of paper. There was a short questionnaire on the back side that asked for age, sex, education, and native language. After reading the problem and providing an answer, subjects were asked to explain in one or two sentences why they decided to turn over or not turn over each of the four items, p , $\sim p$, q , and $\sim q$. Subjects had 15 min to complete the experiment.

Abstract materials. Each of the abstract problems had the same format. We varied only a few key phrases. In the abstract/no-relation problem, subjects were asked to imagine that they were inspecting a large stack of 3 x 5 in. cards with letters (A or B) on one side and numbers (1 or 2) on the other side. The task was to make sure the following rule was obeyed: "If there is an A on one side of the card, then there must be a 2 on the other side." Subjects were specifically instructed that they should minimize the number of cards that they turned over, while at the same time making sure that the rule had been followed. Below these instructions were illustrations of a sample of four cards "selected" from the entire deck ("A," "B," "2," "1"). Left-to-right order of the cards was counterbalanced to eliminate potential order effects.

The abstract-blank condition was the same as the abstract/no-relation condition except that the word "blank" was substituted for the number 1 and a blank card was substituted for the "1" card. Also, the third sentence was changed to read: "Each one of the cards either has a number 2 on the back side or it is blank on the back side." For the abstract-explicit condition, the only difference was the addition of the following phrase: "Cards with the letter A on the front may only have the number 2 on the back, but cards with the letter B on the front may have either 1 or 2 on the back." Because of the stark nature of the abstract materials, implicit relationships between antecedents and consequents were difficult to create. We relied upon the ordinal relationship between the consequent numbers 18 and 5. For this condition, three changes were made: (1) The third sentence read "Each one of the cards has a number on the back side," (2) the rule read "If there is an A on one side of the card, then there must be a number greater than 10 on the other side," and (3) the cards at the bottom showed the letters "A" and "B" and the numbers "18" and "5."

Concrete materials. The concrete materials employed the scenario of a quality-control clerk inspecting different models

of pocket calculators moving along a manufacturing plant conveyor belt. Subjects were told that Microdigit, Inc., markets two different calculator models, the XT-10 and the XT-11; these models were functionally the same, but the XT-10 was sold in the United States and the XT-11 was exported to Canada. Model numbers appeared on the front side and a brief set of instructions could be glued on a panel on the back side. The instructions were printed in two versions, one technical (for the business market) and one quite simple (for the consumer market). The calculators moved past the quality-control clerks on a conveyor belt, some face up with the model number showing and some face down with the instruction panel showing. The clerks had to make sure that the following rule was obeyed: "If a calculator is a Model XT-10, then the simple instructions must be on the panel on the back side." Subjects were told that clerks had to check the calculators as quickly as possible and therefore wanted to minimize the number of calculators they had to turn over while making sure that the rule had been followed in all cases. Below these instructions were illustrations of four calculators ("XT-10," "XT-11," "simple instructions," "technical instructions"). Left-to-right order of the calculators was counterbalanced.

In the concrete-blank condition, instead of being told that calculators had either simple or technical instructions on the back, subjects read the following sentences. "The instructions are quite simple (directed toward the residential consumer market). In some cases, no instructions have been glued onto the panels. These are cases where different language instructions are supplied by the distributors at a later date." The only other change was that on the bottom of the page the calculator showing technical instructions was changed to a calculator with a blank panel. For the concrete-explicit condition, the following phrase was added to the no-relation text right after the subjects were told that the instructions were either technical or simple: "The XT-10 is sold with simple instructions, but the XT-11 is sold either with simple or technical instructions." For the concrete-implicit condition, the following two sentences were substituted for the sentences about simple and technical instructions: "The instructions are printed in two versions. One version is in English and one version is in French." The rule then became: "If the calculator is a Model XT-10, then the English instructions must be on the panel on the back side." "English" and "French" replaced "simple" and "technical" in the illustrations at the bottom of the page. Here, the implicit relationship is based upon subjects' knowledge that Canada is a bilingual country and instructions in either language will be appropriate, whereas in the U.S., English is the only appropriate language.

RESULTS

There were two different sets of response variables: (1) solution of the reasoning task (right or wrong), and (2) subjects' explanations for why they decided to check or not check p , $\sim p$, q , and $\sim q$. Subjects' explanations were coded according to the presence or absence of five classes of inferences, described below. Since both the solution and inference responses were categorical variables, they were amenable to analysis only through multidimensional contingency tables. These qualitative data were analyzed using log-linear models estimated by the method of maximum likelihood (Bock, 1975; Bock & Yates, 1973). The technique is essentially a non-parametric analysis of variance yielding chi-square values (or log-likelihood ratios) instead of F ratios to test the goodness of fit of a particular linear model (Fienberg,

Table 1
Solution Frequencies and Probability of Correct Response

Relation	Solution		P(Correct)
	Number Correct	Number Incorrect	
Abstract Theme			
No Relation	7	18	.28
Blank	12	13	.48
Implicit	15	10	.60
Explicit	14	11	.56
Total	48	52	.48
Concrete Theme			
No Relation	11	14	.44
Blank	19	6	.76
Implicit	19	6	.76
Explicit	20	5	.80
Total	69	31	.69
Overall			
Total	117	83	.59*

*Mean.

1980). The significance of single effects is obtained by examining the difference between the chi-square values for the model that contains the effect and those for the model in which the effect has been deleted, termed χ^2 diff.

Problem Solution

Table 1 shows the number of subjects who solved the problem (checked p and ~q) and who failed to solve the problem for each of the eight relation by theme cells. Across conditions, 59% of all the subjects solved the problem. However, only 28% made the correct response for the modified Wason selection problem, the abstract/no-relation cell.² There also was a noticeable increase in the probability of success for concrete vs. abstract materials, 69% and 48%, respectively. However, the most interesting effect was due to the relation factor. Even in the abstract problems, subjects given explicit or implicit cues performed near or above the overall

mean. The solution percentages for abstract-explicit and abstract-implicit were 56% and 60%, respectively. These subjects' performance was superior to those given the modified Wason problem.

The best-fitting log-linear model for the data in Table 1 contained the main effects of theme and relation with no interaction term; it fit the data extremely well [$\chi^2(3) = .58$, $p = .901$]. The fit of the model could not be significantly improved by adding the Theme by Relation interaction to this two main effects model. Chi-square deletion differences showed that both the main effects were significant [theme, χ^2 diff(1) = 9.91, $p < .005$; relation, χ^2 diff(3) = 15.07, $p < .005$].

Most important, the main effect of relation was present for both abstract and concrete stimuli. The effect of relation for both levels of theme was tested by constructing two linear contrasts: (1) the difference between the no-relation condition and the mean of the other three levels of the relation factor within the abstract theme condition, and (2) the same difference for the concrete theme condition. These two Helmert contrasts allowed us to test the simple main effects (Keppel, 1973) for the relation factor (i.e., the effect of relation at each level of theme). Chi-square deletion differences showed that both simple main effects were significant [relation at abstract, χ^2 diff(1) = 5.5, $p < .025$; relation at concrete, χ^2 diff(2) = 9.24, $p < .005$]. In summary, with the addition of relational cues, performance improved beyond the baseline no-relation condition for both abstract and concrete problems. There was also a significant effect of theme on solution.

Table 2 presents a more detailed breakdown of the subject's selection responses. The matching-bias response, the selection of p and q (Evans & Lynch, 1973), was predominant only in the abstract/no-relation condition. This supports the notion that the matching selection is the response of last resort when the subject is confused about the demands of the task. Others also have found that the matching bias is suppressed when using thematic materials; this has occurred regardless of whether thematic materials led to improved overall performance

Table 2
Frequency of Selection Combinations for Eight Problem Types

Selection	Abstract Problems				Concrete Problems				Total
	No Relation	Blank	Implicit	Explicit	No Relation	Blank	Implicit	Explicit	
Combinations									
p,~q	7	12	15	14	11	19	19	20	117
p	3	5	6	1	6	1	2	1	25
p,q*	10	2	1	1	2	2	2	2	22
p,q,~q	2	0	0	1	0	1	1	0	5
p,~p,q,~q**	0	1	1	0	3	0	0	0	5
Other	3	5	2	8	3	2	1	2	26
Individual Cards									
p	24	21	23	18	23	24	25	25	183
~p	3	4	2	6	4	1	1	2	23
q	12	6	2	7	6	4	3	3	43
~q	10	17	18	16	16	21	20	20	138

*Matching bias. **Biconditional.

(Griggs & Cox, 1982; van Duyne, 1974). Our data show that the matching bias also disappears with abstract materials when information about the relation between the items is included.

Selection Explanations

Our primary interest in the written explanations was to understand why adding relational information improved performance, irrespective of the thematic content of the problem. Evans and Wason (1976) have proposed a dual process theory in which selections and justifications are considered distinct thought processes. They argue that explanations are really rationalizations rather than indicators of the mental processes underlying selection. Although subjects' explanations probably involve some form of rationalization, it does not necessarily follow that rationalization is independent of selection process. Furthermore, judicious analysis of differences in post hoc explanations can be as useful as differences in concurrent protocols to identify different strategic approaches to the task. The explanations can provide evidence convergent with the dichotomous solution measure (Campbell & Fiske, 1959).³

Inference categories. All explanations were categorized into inference classes. We define inferences as statements about the antecedents, consequents, or the rule (p implies q) that did not appear in the problem but were inferred from the information provided in the problem.

p Explanations. Over 91% of the 200 subjects correctly decided to select the p item, with no difference between conditions. Subjects' explanations about p did not discriminate between those subjects who did or did not solve the entire problem. The 160 explanations for p were very similar for most of the subjects. The typical reason for checking p (87% of the explanations) was "to make sure the rule was followed" or "to make sure that q is on the other side." Therefore, we did not consider p explanations in further analyses.

$\sim p$ Explanations. For the $\sim p$ item, there were two predominant inferences that accompanied the correct decision to not check this item. Explanations were coded as " $\sim p$ -irrelevant" if the subject explicitly stated that " $\sim p$ is irrelevant to the rule" or that "the rule does not refer to $\sim p$." Explanations coded as " $\sim p$ implies both q and $\sim q$ " included explanations that ranged from "it doesn't matter which item, q or $\sim q$, is on the back side" to "either q or $\sim q$ are allowed on the back of $\sim p$." While all the subjects who inferred $\sim p$ -irrelevant or $\sim p$ -implies-both correctly did not select $\sim p$, not all of these subjects solved the complete problem. An "other" category consisted of the few explanations that did not fall into the first two categories. This other category included some explanations from subjects who did not select $\sim p$ (e.g., "assume $\sim q$ on the other side") and all explanations of those who erroneously selected $\sim p$ (e.g., "check for presence of q or $\sim q$ ").

q Explanations. For the q item, there was only one predominant inference that accompanied the decision to not check the item. Explanations were coded as " q implies both p and $\sim p$ " when subjects said that "it doesn't matter which item, p or $\sim p$, is on the other side of q " or explicitly stated that " q can be on the back of either p or $\sim p$." The "other" category included explanations of all other subjects, both those who did and did not make the correct decision about q . Examples included " q is irrelevant" (q correctly not selected) and "must check for p on the other side" or "must check because q is part of the rule" (q incorrectly selected).

$\sim q$ Explanations. For the $\sim q$ item, there were two distinct explanations that accompanied the correct decision to check the other side. First, subjects stated that they must check $\sim q$ because "if there is a p on the other side then the rule has been violated." Second, subjects said that they must check $\sim q$ because "only a $\sim p$ can be on the back of $\sim q$." This is equivalent to inferring the contrapositive, " $\sim q$ implies $\sim p$," of the rule " p implies q ." Whether subjects inferred this rela-

Table 3
Proportion of Subjects Providing Explanations Categorized by Inference and Problem Type

Relation	$\sim p$ Inferences				q Inferences			$\sim q$ Inferences		
	N	$\sim p$ Implies Both	$\sim p$ Irrelevant	Other	N	q Implies Both	Other	N	$p, \sim q$ Violates	Other
Abstract Problem										
No Relation	15	.13	.67	.20	17	.29	.71	16	.44	.56
Blank	21	.19	.67	.14	21	.48	.52	21	.71	.29
Implicit	19	.47	.53	.00	21	.81	.19	21	.71	.29
Explicit	19	.58	.32	.11	20	.80	.20	18	.71	.29
Concrete Problem										
No Relation	16	.13	.50	.38	16	.50	.50	16	.62	.38
Blank	19	.37	.32	.32	18	.61	.39	18	.67	.33
Implicit	19	.68	.26	.05	18	.78	.22	20	.85	.15
Explicit	19	.89	.11	.00	19	.95	.05	17	.82	.18
Overall										
Mean	147	.45	.41	.14	150	.66	.34	147	.70	.30

Note— N = the number of subjects (of a possible 25) providing explanations, regardless of whether they made the correct selection decision for that particular item.

tionship by using a formal rule is another question. However, because there were so few of the latter explanations, we combined the two into one category, "p with $\sim q$ violates" the rule. Examples of explanations in the "other" category included "must check $\sim q$ to prove the rule" ($\sim q$ correctly selected) and " $\sim q$ is irrelevant" or "only p and q are part of the rule" ($\sim q$ incorrectly not selected).

Two of the five coded inferences, " $\sim p$ implies both q and $\sim q$ " and "q implies both p and $\sim p$," are deducible from the implication rule "p implies q" within the propositional calculus of material implication. As mentioned earlier, however, subjects typically consider evidential pairs containing $\sim p$ to be irrelevant to the rule (Johnson-Laird & Tagart, 1969), and, in fact, the problem is solvable without consideration of the ($\sim p, q$) and ($\sim p, \sim q$) pairs. Except in the explicit condition, in which subjects were told that $\sim p$ implied both q and $\sim q$ and could easily deduce that q implied both, the two inferences are "extra" logical inferences. Explicit condition subjects were eliminated from our reported analyses of extra logical inferences about $\sim p$, q, and $\sim q$. However, in all cases in which significant results were reported, the effects were identical when explicit subjects were included. Explicit condition results are reported in Tables 3-5 for comparison.

Inference Analysis

The analysis of the inference data involved two steps. In the first set of analyses, inferences about $\sim p$, q, or $\sim q$ served as response variables, with relation and theme serving as independent variables in log-linear models. Table 3 summarizes these data. The second set of analyses examined the differences in overall solution as a function of what inferences the subject made. These data are presented in Table 4 collapsed across problem type. The assumption underlying our analyses was that relational information would influence inference production, and inference production would determine whether or not the subject solved the "overall" problem correctly.

$\sim p$ Inferences. The best-fitting log-linear model contained the two main effects of theme and relation [$\chi^2(4) = .98$, $p = .913$]. However, by examining the

chi-square differences due to deleting effects, the effect of theme was significant only for the $\sim p$ -irrelevant inference [$\chi^2 \text{diff}(1) = 8.47$, $p < .005$]. That is, subjects made significantly more $\sim p$ -irrelevant inferences when given abstract problems (34/55 subjects, 62%) than when given concrete problems (19/54 subjects, 35%). Moreover, the effect of relation was present only for the $\sim p$ -implies-both inference [$\chi^2 \text{diff}(2) = 16.37$, $p < .001$]. Contrasts were constructed that allowed us to examine differences between levels of the relation factor for $\sim p$ -implies-both. Across levels of theme, implicit condition subjects were more likely than either blank or no-relation subjects to infer that $\sim p$ implied both q and $\sim q$, the information stated in the explicit condition.

Subjects who made the $\sim p$ -implies-both inference were more successful in solving the problem than were those who state $\sim p$ -irrelevant or "other" (see Table 4) [$\chi^2(1) = 18.23$, $p < .001$]. This difference was still significant when comparing only $\sim p$ -implies-both and $\sim p$ -irrelevant [$\chi^2(1) = 4.67$, $p < .05$]. One possible clue to the inferior performance of subjects inferring $\sim p$ -irrelevant comes from examining their explanations for $\sim q$. Nineteen of the 61 subjects who inferred $\sim p$ -irrelevant also said " $\sim q$ is irrelevant"; while this conditional probability is not large (.31), none of the subjects who made the $\sim p$ -implies-both inference said " $\sim q$ is irrelevant," a false conclusion. Saying that $\sim p$ and $\sim q$ are irrelevant is akin to the matching bias, symptomatic of a response of last resort.

q Inferences. For the q inferences shown in Table 3, the best-fitting log-linear model contained only the main effect of relation [$\chi^2(2) = 2.25$, $p = .52$]. Again, contrasts were constructed to test differences in inference production for levels within the relation factor. Implicit condition subjects were more likely to infer q-implies-both than were either blank or no-relation subjects in abstract and concrete problems.

For the q inference data in Table 4, subjects making the q-implies-both inference solved the problem significantly more often than those who gave other explanations [$\chi^2(1) = 21.22$, $p < .001$]. Table 5 shows the proportions of subjects who correctly solved the problem and made both of the extra logical inferences, $\sim p$ -implies-both and q-implies-both, either inference, or

Table 4
Number of Subjects Providing Explanations Who Solved or Did Not Solve the Task Categorized by Inference

	$\sim p$ Inferences			q Inferences		$\sim q$ Inferences	
	$\sim p$ Implies Both	$\sim p$ Irrelevant	Other	q Implies Both	Other	p, $\sim q$ Violates	Other
All Problems							
Solvers	57	36	4	82	17	94	5
Nonsolvers	8	25	17	17	34	9	39
Excluding Explicit Problems							
Solvers	31	32	4	52	17	70	3
Nonsolvers	7	21	15	13	29	6	33

Table 5
Numbers of Subjects Providing Extra Logical Explanations
Who Solved or Did Not Solve the Task

	All Problems			Excluding Explicit Problems		
	Both	Either	Neither	Both	Either	Neither
Solvers	53	32	21	29	27	21
Nonsolvers	4	18	33	2	16	28

Note—Both = $\sim p$ -implies-both and q -implies-both; either = $\sim p$ -implies-both or q -implies-both; neither = neither $\sim p$ -implies-both nor q -implies-both.

neither. Subjects who made both inferences were more successful than those who made only one [$\chi^2(1) = 9.26$, $p < .005$].

$\sim q$ Inferences. For the $\sim q$ inference data in Table 3, the best-fitting log-linear model included the two main effects of theme and relation [$\chi^2(2) = 1.34$, $p = .512$]. However, closer examination of the model revealed that neither of these main effects was particularly large, as chi-square differences due to deleting either main effect were not significant. Only in the abstract/no-relation condition did a substantial number of subjects, 44%, fail to say that p with $\sim q$ violates the rule. Generally, the theme and relation factors had little systematic effect upon $\sim q$ inference production.

As can be seen from Table 4, subjects who inferred that p with $\sim q$ violates the rule were much more likely to solve the problem than were subjects making inferences in the other category [$\chi^2(1) = 75.53$, $p < .001$]. However, it is unclear what this really means, since this inference is almost a necessary condition for correct solution. The situation is quite different when it comes to inferences about $\sim p$ and q . While inferring either $\sim p$ -implies-both or q -implies-both is not a necessary condition for solution of the problem, these inferences do seem to be sufficient in many cases, as evidenced by their high conditional probabilities of success, .82 for $\sim p$ -implies-both and .83 for q -implies-both (excluding the explicit conditions).

DISCUSSION

Solution of the selection task requires that subjects seek disconfirming evidence from the sample cards in as efficient a manner as possible. Changes in propositional content from the basic no-relation condition lead to significant increases in solution probabilities. The analysis of the explanation data indicate that subjects can use quite different solution strategies depending upon the nature of the redundant cues provided by the propositional content. One type of cue encourages subjects to construct counterexamples ($p, \sim q$) either from known rules and regulations (Griggs & Cox, 1982; Johnson-Laird et al., 1972) or from more general beliefs induced from a range of experiences (e.g., the presence

of a blank item). A second redundant cue encourages subjects make extra logical inferences about ($\sim p, q$) and ($\sim p, \sim q$). Finally, beyond the redundant interitem relational cues provided by propositional content, additional information is provided by the task scenario or theme.

Extra Logical Inferences

The analysis of the inference data provided evidence for the importance of recognizing allowable as well as illegal antecedent-consequent relations for each of the four sample cards. Although the extra logical inferences $\sim p$ -implies-both and q -implies-both were not necessary to solve the problem, solution probabilities were much higher for those subjects who did make these inferences, especially in the implicit and explicit conditions. These subjects behaved as if they had constructed a complete truth table (i.e., all pairs are true or false, not irrelevant).

The stated implication rule encouraged subjects to consider the front and back sides of the p card and to turn over the card to confirm the rule. The presence of the $\sim p$ -implies-both and q -implies-both inferences indicates that some subjects considered both sides of the $\sim p$ and q card and discovered a reason not to turn over these cards because it did not matter what was on the other side. This inference process might have encouraged these subjects to consider the final card, $\sim q$, thus generating the remaining possible antecedent-consequent pair. Subjects now had all the information implied by the conditional and could solve the problem. We believe that many of the difficulties subjects experience in the selection task reside in the inability to generate all possible antecedent-consequent pairs rather than in correctly identifying a pair's truth value. Goodwin and Wason (1972) made a similar observation: "It is not that subjects cannot perform a combinatorial analysis (generate and identify) with abstract materials, but that the idea of doing so seldom occurs to them" (p. 211). We know that subjects can correctly identify the ($p, \sim q$) pair as false when they see it (Johnson-Laird & Tagart, 1969). The so-called reversibility problem (i.e., checking p for $\sim q$, but not checking $\sim q$ for p) seems to be symptomatic of more general difficulties in generating pairs rather than problems in identifying truth values. The fact that the $\sim q$ -irrelevant inference often accompanied the $\sim p$ -irrelevant inference indicates that some subjects focused only upon those items mentioned in the rule and did not generate all possible pairs. The information about the allowable interitem relations of $\sim p$ and q provided in the implicit and explicit conditions led to more extra logical inferences and an exhaustive generation of all antecedent-consequent pairs.

The explanation data support the above analysis and suggest that many subjects in both concrete and abstract conditions were not relying solely upon domain-specific counterexamples to solve the problem. Consider a comparison between our abstract-implicit problem and the Griggs and Cox (1982) drinking-age task. The drinking-

age rule was "if a person is drinking beer, then the person must be over 19"; the abstract-implicit rule was "if the letter A is on one side, then the number on the other side must be greater than 10." While the rules are analogous, subjects given the abstract-implicit problem had no known civil law to apply to the situation. They had to rely on other inference procedures rather than the production of counterexamples to solve the problem, in this case extra logical inferences about allowable relations for $\sim p$ and q .

In the concrete-implicit problem, there were two different but related cues that subjects could have used. They could have identified a salient counterexample to the rule, in this case an XT-10 calculator, meant for the U.S. market, with French instructions on the back. They also could have relied upon their knowledge that Canada is a bilingual country to infer that the XT-11 can have English or French instructions. The counterexample is probably sufficient to solve the problem, but the high incidence of extra logical inferences suggests that subjects constructed the entire truth table.

Counterexamples

Griggs and Cox (1982) and Manktelow and Evans (1979) postulate that the thematic materials effect is due to long-term memory cuing, in which subjects draw upon domain-specific, experiential knowledge to solve the problem. The counterexamples serve as existence proofs, cuing subjects to the fact that rule violations may occur and that potentially illegal combinations must be checked. Counterexamples to a rule may encourage subjects to adopt a "detective set" (van Duyne, 1974) rather than to act as mere rule enforcers, as they do in abstract/no-relation problems. Our results show that performance also improves even when cues to illegal pairs of a more general nature are provided, as evidenced by the blank conditions.

The explanation data showed that subjects who generated extra logical inferences ($\sim p$ -implies-both and q -implies-both) solved the problem more often than those subjects who did not. However, this relationship was not nearly as pervasive in the blank conditions, in which these extra logical inferences were generated less frequently. However, the probabilities of solving blank, implicit, and explicit problems were almost identical for both abstract and concrete versions. Blank problems did not appear to encourage exhaustive enumeration of the allowable antecedent-consequent pairs. Subjects were able to dismiss the ($\sim p, q$) and ($\sim p, \sim q$) pairs as irrelevant and still solve the problem. We attribute this behavior to what we call the "red-flag" phenomenon. Past experience with situations in which something was missing might have taught subjects to be wary and take extra care. Absence was an easily recognized cue that immediately distinguished the blank object from those that were "normal." While subjects might not have known what was wrong, they did know that something was wrong.

Concrete Content

While we can only speculate on the basis for the superiority of one theme over another, a plausible conjecture involves what Wason and Shapiro (1971) called the "ease of manipulating" concrete symbols. There is abundant evidence that concrete items are better remembered than abstract ones (Paivio, 1971). Mental operations on representations of tangible object items could require less effort and be less prone to error. We have hypothesized that one solution strategy involves the generation of possible antecedent-consequent pairs and the subsequent identification of "illegal" pairs. The solution and inference data provide evidence that performance also improves when subjects recognize all the allowable pairs. While exhaustive enumeration of all possible antecedent-consequent pairs improves performance, it also places a burden on short-term memory. Subjects must coordinate eight antecedent-consequent pair representations. The prevalence of " $\sim p$ -irrelevant" inferences in abstract problems suggests that subjects may have more difficulty pairing off abstract items and maintaining them in memory. With abstract propositions, subjects rely upon the $\sim p$ -irrelevant inference as a rationale for not selecting $\sim p$. While this is often pragmatically valid, it is a logically false inference and symptomatic of other possible defects in subjects' mental representation of the information in the truth table.

Summary

In summary, manipulating the content of reasoning problems has dramatic effects upon performance. While this is not a unique observation, our research clarifies the basis for the thematic materials effect. Our subjects were skillful in capitalizing upon a variety of cues that functioned vicariously with the logical structure of the task. They did not rely solely upon retrieval of domain-specific knowledge from long-term memory to generate counterexamples. Although different cues may result in different solution strategies, changes in content that provide redundant cues increase the probability that subjects will behave in a way consistent with the logical model. The four-card selection task frequently has been cited as evidence of the frailty of human inference processes. Our subjects, however, demonstrated an ability to coordinate extraexperimental knowledge in addition to the implication rule to reason in a logical fashion.

REFERENCE NOTE

1. D'Andrade, R. *Reason versus logic*. Paper presented at the Symposium on the Ecology of Cognition: Biological, Cultural, and Historical Perspectives, Greensboro, North Carolina, April 1982.

REFERENCES

- Bock, R. D. Multivariate analysis of qualitative data. In R. D.

- Bock (Ed.), *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill, 1975.
- BOCK, R. D., & YATES, G. *MULTIQUAL: Log-linear analysis of nominal or ordinal qualitative data by method of maximum likelihood*. National Resources, 1973.
- BRUNSWIK, E. The conceptual framework of psychology. In *International encyclopedia of unified science* (Vol. 1, No. 10). Chicago: University of Chicago Press, 1952.
- CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the multi-trait, multi-method matrix. *Psychological Bulletin*, 1959, **56**, 159-176.
- EVANS, J. ST. B. T., & LYNCH, J. S. Matching bias in the selection task. *British Journal of Psychology*, 1973, **64**, 391-397.
- EVANS, J. ST. B. T., & WASON, P. C. Rationalization in a reasoning task. *British Journal of Psychology*, 1976, **67**, 479-486.
- FIENBERG, S. E. *Analysis of cross-classified categorical data*. Cambridge, Mass: M.I.T. Press, 1980.
- GOODWIN, R. Q., & WASON, P. C. Degrees of insight. *British Journal of Psychology*, 1972, **63**, 205-212.
- GRIGGS, R. A., & COX, J. R. The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, 1982, **73**, 407-420.
- HAMMOND, K. R. 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, 1966.
- JOHNSON-LAIRD, P. N., LEGRENZI, P., & SONINO-LEGRENZI, M. Reasoning and sense of reality. *British Journal of Psychology*, 1972, **63**, 395-400.
- JOHNSON-LAIRD, P. N., & TAGART, J. How implication is understood. *American Journal of Psychology*, 1969, **82**, 367-373.
- JOHNSON-LAIRD, P. N., & WASON, P. C. A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, 1970, **1**, 134-148.
- KEPPEL, G. *Design and analysis: A researcher's handbook*. Englewood Cliffs, N.J: Prentice-Hall, 1973.
- MANKTELOW, K. I., & EVANS, J. ST. B. T. Facilitation of reasoning by realism: Effect or non-effect. *British Journal of Psychology*, 1979, **70**, 477-488.
- PAIVIO, A. U. *Imagery and verbal processes*. New York: Holt, Rinehart & Winston, 1971.
- VAN DUYNE, P. C. Realism and linguistic complexity in reasoning. *British Journal of Psychology*, 1974, **65**, 59-67.
- VAN DUYNE, P. C. Necessity and contingency in reasoning. *Acta Psychologica*, 1976, **40**, 85-101.
- WASON, P. C. Reasoning. In B. Foss (Ed.), *New horizons in psychology*. London: Penguin, 1966.
- WASON, P. C., & JOHNSON-LAIRD, P. N. *Psychology of reasoning: Structure and content*. Cambridge, Mass: Harvard University Press, 1972.
- WASON, P. C., & SHAPIRO, D. Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 63-71.
- YACHANIN, S. A., & TWENEY, R. D. The effect of thematic content on cognitive strategies in the four-card selection task. *Bulletin of the Psychonomic Society*, 1982, **19**, 87-90.

NOTES

1. Yachanin and Tweney (1982) argue that the modified task is easier because only one hypothesis (the rule) need be processed rather than two (true or false rule), thus requiring different psychological processes. Another possibility is that the original instructions simply are more ambiguous. In developing our stimuli, we tried to be as clear as possible so that performance differences between problems could not be attributed to total confusion about the demands of the abstract task. At any rate, our modified version of the problem, which asks subjects to make sure a rule is followed, seems at least intermediate in difficulty when compared to those cases in which subjects are asked to ensure that a rule has not been violated (e.g., Griggs & Cox, 1982), which seems to encourage falsification.
2. The 28% solution rate is quite a bit higher than the typical 10% rate for abstract/no-relation problems. One reason might be due to population differences that have been observed between subjects in Great Britain, Australia, and the United States, possibly related to intelligence and education level. A second reason may be due to the use of the modified task (Yachanin & Tweney, 1982).
3. A potential limitation to these introspective data concerns a selection bias due to missing data. Of 200 subjects, there were 135 cases with explanations for all four items, 26 cases with explanations for some of the items, and 39 cases with no explanations at all. Of the 117 subjects who solved the problem correctly, 90.5% provided some usable explanations; of the 83 nonsolvers, only 66.3% provided some explanations [$\chi^2(1) = 18.31$, $p < .001$]. Failure to provide explanations was due to many factors, including a lack of time (15 min to complete the task), lack of motivation, and failure to follow the complete set of instructions. Because of these differences, the two groups of subjects, solvers and nonsolvers, may not be comparable.

(Received for publication April 9, 1982;
revision accepted August 23, 1982.)