

The effect of thematic content on cognitive strategies in the four-card selection task

STEPHEN A. YACHANIN and RYAN D. TWENEY
Bowling Green State University, Bowling Green, Ohio 43403

Prior studies have indicated that thematic materials facilitate reasoning on the four-card selection problem, a task that assesses ability to evaluate logical conditionals of the "if-then" form. Recently, however, attempts to replicate the thematic effect have failed. The present study investigated four content materials to determine which facilitated reasoning. Introductory psychology students were required to determine which card(s) needed to be "unmasked" in order to determine the truth or falsity of conditional rules. Thematic materials did not facilitate reasoning compared with abstract materials. Evidence for both matching and verification bias was obtained, but neither was sufficient to account for all selection patterns. Subjects may thus be using "cognitive short-circuiting" strategies, of which matching bias and verification bias are but two examples.

Knowledge of the conditions under which hypotheses about the world are combined with evidence to reach a true or false judgment is a central requirement for a complete theory of human reasoning. It has been known for some time that, logically, evidence that confirms a hypothesis does not prove the truth of the hypothesis. On the other hand, evidence that falsifies a hypothesis can lead to certainty: certainty that the hypothesis is false. Thus, falsification is often prescribed as the method of favor for testing hypotheses (Platt, 1964; Popper, 1963). In a seminal series of studies, Wason (1968) used a particular verification procedure, the four-card selection task, to investigate how people use falsification. In one form of the experiment, four cards with a letter and a number on each side were presented to a subject. Either a letter or a number was visible on each card (e.g., E, K, 4, or 7). The other element was always covered by a mask. A conditional rule "If a card has an E on one side, then it has a 4 on the other side" (if P then Q) was given to the subject. The subject's task was to select only those cards that needed to be unmasked in order to determine if the rule was true or false. Wason found that two consistent errors were made by most subjects: pervasive selection of the Q card and failure to select the $\neg Q$ (not Q) card (Wason, 1968; Wason & Johnson-Laird, 1970). In fact, only the combination P, $\neg Q$ can possibly falsify the rule; all other combinations are consistent with either the truth or falsity of the rule and have no inferential value. Selection of the combination P,Q (the true antecedent, TA, and the true consequent, TC) and failure to recognize the importance of the falsifying case $\neg Q$ (the false

consequent, FC) was called verification bias (Johnson-Laird & Wason, 1970).

Using the same task, Evans and Lynch (1973) conducted an experiment that suggested subjects chose only those cards named in the conditional, a process they called matching bias. In the traditional selection task, only one rule had been tested. Such a procedure confounds matching and verification bias, since the combination P,Q is both the verifying and the matching response. Evans and Lynch showed that it is possible to separate matching and verification bias if negative qualifiers of the propositions are introduced. Table 1 shows the four rules possible when negative qualifiers are included. Subjects with a verification bias should choose the TA only or the TA and TC cards, regardless of the presence of qualifiers, whereas the asterisked cards should be chosen by subjects with a matching bias. Evans and Lynch found that the card selections of most subjects were consistent with a matching bias and not generally with a verification bias.

Wason and Shapiro (1971) found that it was possible to circumvent verification bias if thematic materials were used as content materials for the rules. They pre-

Table 1
Affirmed (Matching) and Negated (Mismatching) Values
Constituting the Four Logical Choices on Each
Form of the Conditional Rules

Rule	Logical Case			
	TA	FA	TC	FC
(1) If P then Q	P*	$\neg P$	Q*	$\neg Q$
(2) If P then not Q	P*	$\neg P$	$\neg Q$	Q*
(3) If not P then Q	$\neg P$	P*	Q*	$\neg Q$
(4) If not P then not Q	$\neg P$	P*	$\neg Q$	Q*

Note—Adapted from Evans and Lynch (1973, p. 392).

*Matching values.

The authors would like to thank Michael E. Doherty and Irwin W. Silverman for their helpful comments on the conduct of this study, and Peter Wason and Richard Griggs for advice and discussions of the four-card task.

sented subjects with conditional rules set in the thematic guise of a journey to a town by a specific mode of transportation: "Every time I go to Manchester I travel by car." Ten of 16 subjects selected the correctly falsifying combination TA,FC in the thematic material condition, compared with only 2 of 16 subjects in a control condition using letters and numbers. Wason and Shapiro argued that a thematic rule formed a coherent whole that permitted subjects to attend equally to each card. Thus, subjects were able to generate alternative values for the hidden sides of the cards and to evaluate the truth or falsity of the rule with each combination. Wason and Shapiro argued that when abstract stimuli were used, subjects resorted to matching the named values in a one-to-one fashion.

If thematic content facilitates correct reasoning, then matching bias should not be observed when negative qualifiers are introduced in thematic rules. Manktelow and Evans (1979) tested this in four experiments, but they found no difference between abstract and thematic material conditions; matching was observed in both conditions. Further, they failed to replicate the original Wason and Shapiro (1971) finding in a fifth experiment, in which the circumstances of the original Wason and Shapiro study were closely recreated. Pollard (in press) did succeed in replicating the Wason and Shapiro finding. He argued that the thematic content used in the first four experiments by Manktelow and Evans ("foods and beverages") cannot be assumed to be a sample of the universe of thematic content and may actually have been "abstract," in that there was nothing inherently realistic about the relationships used (Van Duyne, 1974). Pollard suggested that differences in subject population and power limitations due to small samples could explain failures to replicate the effect.

Griggs and Cox (in press) also failed to replicate the thematic effect using a procedure similar to that of Wason and Shapiro (1971). They further tested the procedure of Johnson-Laird, Legrenzi, and Legrenzi (1972), in which a thematic effect was found by having subjects determine if a rule of the form "If a letter is sealed, then it has a 50 lire stamp on it" was being violated. Griggs and Cox failed to obtain the thematic effect with this form of the task, but they succeeded when the rule was consistent with their subjects' experience. Thus, subjects did correctly seek falsification for rules of the form "If a person is drinking beer, then the person must be over 19 years of age," although there was no such effect for rules similar to those used by Johnson-Laird et al. (1972). However, it is clear that these tasks differ in an important way from the traditional four-card task. Instead of determining the truth or falsity of the rule, subjects can take the rule as given and evaluate the consistency of the instances. The task is easier, since only one hypothesis, the rule as stated, must be processed, rather than two: that the rule is true or that the rule is false. Thus, the Griggs and Cox and Johnson-Laird et al.

studies do not establish whether a thematic effect exists in true inferential tasks.

If ability to reason in a logical manner is, in part, a function of content, it is important to know what materials enable subjects to correctly test logical rules. However, no systematic investigation of content has been conducted using a single population. The present study accordingly compared four different contents used in previous studies. Negative qualifiers were systematically varied to determine if matching bias varied as a function of thematic content (assuming that a thematic effect could, in fact, be demonstrated).

METHOD

Subjects

One-hundred and sixty Bowling Green State University undergraduates served as subjects (114 females, mean age = 18.4 years; 46 males, mean age = 18.7 years). Participation fulfilled an introductory psychology course requirement.

Design and Materials

Subjects were tested in one of four content conditions. Each subject was asked to test the truth or falsity of two examples of each rule form in Table 1. Rule forms were counterbalanced to yield four rule-order presentations.

The selection task was presented to each subject in the form of a booklet, consisting of two pages of typed instructions and eight test sheets (two for each rule form). Letters and numbers were used for "letter" rules, foods and beverages for "food" rules, modes of transportation and Bowling Green locations for "location" rules, and schools and major fields of study for "school" rules. All combinations in the food, location, and school rules were expected to be consistent with the experiences of the subject population.

Procedure

All subjects were tested in a single group and were assigned to conditions randomly. Subjects were required to test the truth or falsity of the eight conditional sentences by placing a check mark below the card(s) they needed to unmask and an X below the card(s) they did not need to unmask to test the truth or falsity of each rule.

RESULTS AND DISCUSSION

Ninety-eight of 160 subjects (61%) correctly selected the TA/FC card combination for at least one rule, a percentage close to the 62% observed by Wason and Shapiro (1971), suggesting that subjects were performing at a fairly high level in all conditions. However, many subjects selected TA/FC only for the "If P then not Q" rule, in which P and Q are the matching responses. When such selections were not included, performance dropped to 22%. Although they performed at a higher level than the 7% reported by Wason (1968) for abstract content, most subjects still responded in a logically nonoptimal fashion.

Table 2 contains the mean number of TA/FC card selections (i.e., correctly falsifying selections) made by subjects in each content material condition for each rule. Subjects in the food condition showed the best performance, but subjects in all conditions performed

poorly; there were no significant differences between contents.

Table 3 contains the frequencies with which subjects chose each alternative for each rule. One-tailed sign tests were conducted for each matching-bias hypothesis. More TA cards were selected for rules with affirmative antecedents ($z = 5.20, p < .01$). More FA cards were selected for rules with negative antecedents ($z = 6.00, p < .01$). More TC cards were selected for rules with affirmative consequents ($z = 5.54, p < .01$). And more FC cards were selected for rules with negative consequents ($z = 6.08, p < .01$). Thus, each matching-bias hypothesis was confirmed.

The rank order of selections pooled over all rules was $TA > TC > FC > FA$, which suggests that, in spite of matching-bias effects, verifying responses were most frequently selected. The same order was reported by Manktelow and Evans (1979). Two-tailed sign tests revealed that TA was selected significantly more often than TC, FA, and FC ($p < .01$ in all cases). The TC and FC cards were also selected significantly more often than the FA card ($p < .01$ in both cases). A two-tailed sign test revealed that TC was selected significantly more often than FC ($p < .01$), as predicted by verification bias.

The matching-bias hypothesis predicts overall preference for TA/TC for "If P then Q," TA/FC for "If P then not Q," FA/TC for "If not P then Q," and FA/FC for "If not P then not Q." The verification-bias hypothesis predicts that subjects with no insight into the conditional rule should select only TA and TC, or TA alone.

Table 4 contains the frequencies with which card

Table 4
Frequencies With Which Subjects Chose All Possible Card Combinations

Card Combination	P,Q	P,-Q	-P,Q	-P,-Q
TA Only	47	62	13	6
FA Only	1	3	10	12
TC Only	7	6	40	4
FC Only	1	4	0	18
TA, FA	1	1	1	17
TA, TC	161†	41	66	89
TA, FC*	5	116†	23	19
FA, TC	1	21	83†	5
FA, FC	54	2	10	71†
TC, FC	3	0	2	5
TA, FA, TC	3	7	18	6
TA, FA, FC	3	5	4	15
TA, TC, FC	15	19	15	13
FA, TC, FC	8	5	14	3
All Cards	3	25	16	27
No Cards	7	3	5	10

Note— $N = 160$; each subject tested each rule twice. Therefore, columns sum to 320. *Correct response for all rule forms. †Matching response.

combinations were chosen for each rule.¹ Entries were obtained by pooling across content. Inspection reveals that matching-bias selections were most frequent for all rules except "If not P then not Q," in which the matching combination was the second most frequent response. The verifying combination TA/TC was also frequently chosen. However, frequent selection of TA/TC could also be a higher order matching strategy. That is, if subjects found it cognitively difficult to deal with negations, they could have reformulated the negated propositions into nonnegated forms. If so, then TA/TC is a matching response. This hypothesis is consistent with the high observed frequency of TA/TC selections. Further, the higher frequency of FA/FC for "If P then Q" may be the result of a contorted disconfirmation strategy in which the false cases are selected for the antecedent and consequent of the proposition. Finally, it is apparent from Table 4 that strategies other than verification or matching must have been used by at least some subjects. Thus, for "If P then Q," TA alone was selected 47 times, and for "If P then not Q," TA alone was selected 62 times. For the remaining two rules, however, TA alone was seldom selected (13 and 6 times, respectively). Similarly, TC alone was selected 40 times for "If not P then Q," and FA and FC were selected 54 times for "If P then Q."

Why does such a variety of strategies exist? One possibility is that the extra demands of formal reasoning may result in a search for a "cognitive short-circuiting" strategy, a heuristic, to reduce the load. By using such a method, a subject can respond consistently to the task while expending a minimum of cognitive effort. Doherty, Mynatt, Tweney, and Schiavo (1979) conducted a study that suggested that subjects use such cognitive load-reducing strategies. A tendency was observed for subjects to select pseudodiagnostic evi-

Table 2
Mean TA/FC Card Selections for the Four Content Materials Used in the Experiment

	P,Q	P,-Q	-P,Q	-P,-Q	All Rules
Letter	.000	.575	.175	.150	.90
Food	.075	.775	.125	.175	1.15
Location	.025	.875	.100	.025	1.03
School	.025	.675	.175	.125	1.00
Column Average	.031	.725	.144	.119	1.02

Note— $N = 40$ per row. Maximum cell entry = 2.00 for Columns 1-4 and 8.00 for Column 5.

Table 3
Frequencies With Which Subjects Chose Each Logical Alternative for Each Rule

Rule	Logical Case			
	TA	FA	TC	FC
(1) If P then Q	238	75	200	92
(2) If P then not Q	276	69	125	177
(3) If not P then Q	155	156	253	85
(4) If not P then not Q	190	151	157	176
Overall Percent	67.1	35.2	57.4	41.4

Note—Maximum cell entry = 320.

dence, that is, evidence irrelevant to the solution of an inference task in which they sought information to confirm a probabilistic diagnosis. Subjects apparently found it easier to confirm a single hypothesis than to test several alternative hypotheses, even when relative effort was equated. In the same way, in evaluating evidence in the four-card task, it may be more difficult to test two hypotheses, that the rule might be either true or false, instead of one, that the rule is true.

Matching bias may also be considered a cognitive short-circuiting strategy. If subjects find testing even one hypothesis too difficult, a matching strategy can result in systematic responses that require only a minimum effort. The reasoning process is short-circuited and the cognitive load is thereby reduced. Griggs and Cox (in press) suggested that the thematic effect will be observed whenever the rule content cues relevant memories from the subjects' experience. This, too, is "short-circuiting." Similarly, Pollard and Evans (1981) suggested that prior experience with a particular content leads subjects to pick cards with the largest number of associative cues.

What remains unclear is the conditions under which such nonlogical devices are used. It may be that logical strategies require that subjects adopt a "detective set"; that is, they adopt a role in evaluating presented evidence that requires consideration of both confirming and falsifying outcomes (Lunzer, Harrison, & Davey, 1972; Van Duyne, 1976). It is not presently clear whether such a set can be readily induced with abstract materials. Exploration of this possibility will clarify the nature of the heuristics that underlie evaluation of evidence, and such exploration is a high priority for future studies of the psychology of reasoning.

REFERENCES

- DOHERTY, M. E., MYNATT, C. R., TWENEY, R. D., & SCHIAVO, M. D. Pseudodiagnosticity. *Acta Psychologica*, 1979, **43**, 111-121.
- EVANS, J. S. B. T. Toward a statistical theory of reasoning. *Quarterly Journal of Experimental Psychology*, 1977, **29**, 621-634.
- EVANS, J. S. B. T., & LYNCH, L. S. Matching bias in the selection task. *British Journal of Psychology*, 1973, **64**, 391-397.
- GRIGGS, R. A., & COX, J. R. The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, in press.
- JOHNSON-LAIRD, P. N., LEGRENZI, P., & LEGRENZI, M. S. Reasoning and a sense of reality. *British Journal of Psychology*, 1972, **63**, 395-400.
- JOHNSON-LAIRD, P. N., & WASON, P. C. A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, 1970, **1**, 134-138.
- LUNZER, E. A., HARRISON, C., & DAVEY, M. The four-card problem and the generality of formal reasoning. *Quarterly Journal of Experimental Psychology*, 1972, **24**, 326-339.
- MANKTELOW, K. I., & EVANS, J. S. B. T. Facilitation of reasoning by realism: Effect or non-effect? *British Journal of Psychology*, 1979, **70**, 477-488.
- PLATT, J. R. Strong inference. *Science*, 1964, **146**, 347-353.
- POLLARD, P. The effect of thematic content on the "Wason selection task." *Current Psychological Research*, in press.
- POLLARD, P., & EVANS, J. S. B. T. The effects of prior beliefs in reasoning: An associational interpretation. *British Journal of Psychology*, 1981, **72**, 73-81.
- POPPER, K. *Conjectures and refutations*. London: Routledge & Kegan Paul, 1963.
- VAN DUYN, P. C. Realism and linguistic complexity in reasoning. *British Journal of Psychology*, 1974, **65**, 59-67.
- VAN DUYN, P. C. Necessity and contingency in reasoning. *Acta Psychologica*, 1976, **40**, 85-101.
- WASON, P. C. Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 273-281.
- WASON, P. C., & JOHNSON-LAIRD, P. N. A conflict between selecting and evaluating information in an inferential task. *British Journal of Psychology*, 1970, **61**, 509-515.
- WASON, P. C., & SHAPIRO, D. Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 63-71.

NOTE

1. Evans (1977) argued that analysis of combinations of choices were of little value, basing his conclusion on the statistical independence of card selections in a variety of four-card studies. However, the choices of Q and -Q were not statistically independent in our data. Thus, for the "If P then Q" rule, subjects chose Q and -Q 29 times, Q alone 172 times, -Q alone 63 times, and neither Q nor -Q 56 times [$\chi^2(1) = 54.17$, $p < .005$].

(Received for publication October 27, 1981.)