British Journal of Psychology (1982), **73**, 407–420 *Printed in Great Britain*

The elusive thematic-materials effect in Wason's selection task

Richard A. Griggs and **James R. Cox**

A series of three experiments was conducted to examine the possible facilitating effect of thematic materials in Wason's selection task. The first two experiments attempted to replicate the major studies indicating such an effect. Experiment 1 failed to replicate the findings of Wason & Shapiro (1971), and the results of Johnson-Laird *et al.* (1972) were not replicated in Expt 2. In support of a memory-cueing hypothesis, improved performance was found in Expt 3 for an implication rule that was a part of our subjects' past experience. The present results are discussed both with respect to the other studies concerning the effect of thematic content in this task and in the context of a memory-cueing explanation for the thematic-materials effect.

This study is concerned with Wason's four-card selection task (Wason, 1966, 1968), The task involves an implication rule usually in the form 'If p then q' and an array of four cards representing p, $\bar{p}$ (not-p), q, and $\bar{q}$ (not-q). In the *basic form* of the task, symbolic materials and affirmative antecedent and consequent are employed. For example, the subject might be presented with four cards showing, respectively, A, 4, D, and 7, and the rule, 'If there is a vowel on one side of a card, then there is an even number on the other side'. The subject is told that each card has a letter on one side and a number on the other side and asked to select just those cards which it is necessary to turn over in order to find out whether the rule is true or false. The correct answer is p and $\bar{q}$ (A and 7 in the example) since only the combination of p and $\bar{q}$ can falsify the rule.

In its basic form, the task is extremely difficult for even highly intelligent subjects. Usually less than 10 per cent of the subjects produce the correct solution [see Wason & Johnson-Laird (1972) for a review of the early literature on this task and Evans (1978) and Johnson-Laird & Wason (1977) for more recent reviews]. Subjects make two basic errors: (1) although most select the p-card, they fail to select the $\bar{q}$-card (failure to apply *modus tollens*) and (2) frequently select the q-card instead (affirming the consequent).

Two studies (Wason & Shapiro, 1971/Expt 2; and Johnson-Laird *et al.*, 1972), however, seem to provide convincing evidence that performance in this task can be facilitated by the use of thematic materials. Wason & Shapiro used rules about four journeys made by the experimenter and involving cities and modes of transportation, such as 'Every time I go to Manchester, I travel by car' (the *universal* form of the implication rule is frequently used). Johnson-Laird *et al.* asked subjects to pretend they were postal workers and to determine whether rules such as 'If a letter is sealed, then it has a 50 lire stamp on it' were violated. For such thematic material, Wason & Shapiro observed a correct response rate of 10/16 (62·5 per cent) and Johnson-Laird *et al.*, 39/48 (81 per cent). This improvement in correct responding will be referred to as the *thematic-materials effect*.

Recently, however, Manktelow & Evans (1979) have questioned the reliability of this effect. In a series of five experiments comparing the use of symbolic and thematic materials in Wason's selection task, no facilitation of performance on thematic rules was observed. Their Expt 5 was an exact replication of the Wason & Shapiro experiment. In addition, Yachanin (1980) has completely failed to find a thematic-materials effect.* This failure is even more dramatic than that of Manktelow & Evans. For the if p then q conditions in Yachanin's experiment, he observed only five correct selections for 240 problems using thematic content—locations and modes of transportation, foods and beverages, and

* The experiments of Yachanin (1980), Pollard (1981), and the present study were all conducted independently of one another with none of the researchers knowing about the other's work until their own work was completed.

0007-1269/82/030407-14 $02.00/0

schools and major fields of study (80 problems of each type). These results argue strongly against Pollard's (1981) hypothesis that Manktelow & Evans's results might be due to a Type II error.

Two other Wason & Shapiro replication studies (Gilhooly & Falconer, 1974; and Pollard, 1981) have found some improvement in performance for thematic materials. These experimenters found *significant* effects due to materials but a fairly low level of correct responding for thematic materials (22 per cent correct by Gilhooly & Falconer and 33 per cent correct by Pollard). One other selection-task study (Lunzer *et al.*, 1972) does show a weak effect of materials. However, as pointed out by Manktelow & Evans (1979), this weak effect of realistic materials is only demonstrated in Lunzer *et al.*'s Expt 1 when combined with a 'reduced' presentation (Insight 2 data), is not apparent at all in their Expt 2, and the methodology used in their Expt 3 is too different from that normally used in selection-task studies to check for a materials effect.*

Two other studies have been fairly successful in replicating Wason & Shapiro's results (Bracewell & Hidi, 1974; and Van Duyne, 1974). In the two conditions of the Bracewell & Hidi study comparable to the two in Wason & Shapiro's experiment, they observed 9/12 (75 per cent) correct responses for the concrete material/natural relationship problems and only 1/12 (8·33 per cent) correct responses for the abstract material/arbitrary relationship problems. One procedural difference in the Bracewell & Hidi study, however, may have at least partially contributed to the good performance. During the instructional phase of the experiment, subjects were told that 'the conditional rule was not reversible'. The overall effect of this instruction seemed to be that a smaller percentage of subjects *in all conditions* included the *q*-card in their responses. Only 11/96 subjects included *q* in their response, while no one in the basic form condition included *q* in their response! The *p* and *q* combination is always one of the most frequent selections for the basic version of the task. Thus, the instruction about non-reversibility of the conditional may have interacted with content and other task and subject factors to bring about the good performance in the conditions of interest. In further support of such a hypothesis, when Bracewell & Hidi reversed the order of the clauses in the implication rule, the thematic-materials effect completely disappeared (only 2 of 12 subjects correctly responded in the concrete material/natural relationship condition).

Van Duyne (1974) examined both the use of thematic content and the linguistic form of the implication rule. The thematic material consisted of student cards taken from a hypothetical register of students, and the rules involved the students' universities and their fields of study. For the two forms of implication normally employed in selection-task studies (the if–then and universal forms), the correct selection rate for thematic material was 14/24 (58·3 per cent) for the universal form and 12/24 (50 per cent) for the if–then form. These rates are comparable to the 62·5 per cent correct performance observed by Wason & Shapiro (1971). The correct response rates for the comparable abstract conditions in Van Duyne's study were below 10 per cent. No difference, however, between correct response rates for thematic and abstract materials was observed for the other two linguistic forms of implication (disjunction and conjunction). In addition, Yachanin (1980) included thematic rules about students' schools and their major fields of study in his selection-task experiment and completely failed to find any thematic-materials effect for such rules.

To the best of our knowledge, there are no *published* replications of the Johnson-Laird *et al.* (1972) results, and these findings provide the most striking evidence for the

* It should be pointed out that Lunzer (1975) argues that Lunzer *et al.* (1972) found 'appreciable and highly significant increases in the percentage of correct solutions when the arbitrary association was replaced... by a realistic image' (p. 373). This is clearly not the case, however, if a correct solution is defined as a complete insight (Insight 2) response (i.e. *p* and $\bar{q}$). Lunzer must be referring to the partial insight (Insight 1) data and/or the data for reduced presentation arrays.

thematic-materials effect. Van Duyne (1976) and Wason (1977), however, do briefly describe an unpublished study by Van Duyne that is very relevant to the Johnson-Laird *et al.* study. Van Duyne not only varied the abstract–thematic dimension but also the arbitrariness of the thematic rules in this unpublished study. As in Johnson-Laird *et al.* (1972), Van Duyne's thematic rules were postal in nature (e.g. the arbitrary rule, 'If there is L.B. MILL on one side of the envelope, then there is PRINTED PAPER REDUCED RATE on the other side'; the non-arbitrary rule, 'If there is PRINTED PAPER REDUCED RATE on one side of the envelope, then it must be left open'). Van Duyne found the arbitrariness of the rule to be the main factor leading to high levels of performance (86·98 per cent and 97·92 per cent correct in the two non-arbitrary, thematic conditions). However, even in the arbitrary, thematic condition, almost 50 per cent correct performance was found.

Given the inconsistency of the replication results for the Wason & Shapiro findings and only Van Duyne's unpublished study for the Johnson-Laird *et al.* results, we decided to attempt to replicate both these studies. In addition, as pointed out by Pollard (1981), temporal factors and population or geographical variables may affect the results for this task. If anything, recent attempted replications (Manktelow & Evans, 1979; Yachanin, 1980; Pollard, 1981) have not been very successful, and very few studies on the selection task have been conducted with American subjects. Thus, our experiments (especially the attempted replication of the Johnson-Laird *et al.* study) seemed necessary in deciding whether the thematic-materials effect is indeed a genuine effect.

Experiment 1 in the present series was a replication of the Wason & Shapiro study, while Expt 2 was a replication of the Johnson-Laird *et al.* study. Experiment 3 examined a long-term memory cueing explanation for improved performance in the selection task with thematic materials (cf. Manktelow & Evans, 1979).

Experiment 1

The subjects in Expt 1 received two selection-task problems (one thematic problem and one abstract problem) instead of just one problem (either thematic or abstract) as in Wason & Shapiro (1971/Expt 2). No strong evidence exists for transfer effects in the selection task with complete presentation arrays. Even with their finding of excellent performance on thematic problems, Johnson-Laird *et al.* (1972) found a striking lack of any form of transfer between thematic and abstract problems. Thus, by having each subject serve as his/her own control, we not only replicated the Wason & Shapiro study on the first trial but collected twice as much data relevant to the thematic-materials effect question by means of our second trial. In addition, the second trial allowed us to check for positive transfer effects in case a facilitative thematic-materials effect was found. The only other methodological change was that we employed American cities rather than British cities and car and plane modes of transportation instead of car and train modes of transportation in the thematic problems. These minor changes were made in order to make the problems more appropriate for our subject population.

Method

Subjects. Thirty-two undergraduates at the University of Florida participated as part of the requirements for a course in introductory psychology.

Design. Subjects were randomly assigned to one of two groups with the constraint that 16 subjects serve in each group. Each subject served as his/her own control and completed a problem with thematic content and a problem with abstract content. One group of subjects was given the thematic-content problem first followed by the abstract-content problem while the other group received the problems in reverse order.

Materials and procedure. Each subject was run individually. In the thematic condition, subjects were given a familiarization deck of 16 cards. The cards were 7·7 × 12·8 cm note cards with the name of a major American city on one side and a mode of transportation on the other side. One of four modes of transportation (car, bus, plane, train) was randomly assigned to each of 16 cities. The subjects were told, 'Here are some cards, please have a look through them. On each card there is a city on one side and a mode of transportation on the other'.

After the familiarization deck was inspected and returned to the experimenter, four cards labelled respectively, Miami, Atlanta, plane, and car, were presented along with a rule specifying the relationship between a city and a mode of transportation. (These four cards were identical to four of the cards in the familiarization deck.) The following instructions were read.

> I will now present you with four cards; the test you are about to do concerns only these four cards. These cards show cities and modes of transportation for four journeys that I took on four different days, with the city to which I traveled on one side and the mode of transportation on the other side. I am going to make a claim about a city to which I travel and the mode of transportation I use, and your task is to say which of the cards would need to be turned over to decide whether that claim is true or false. You may choose any or all of the cards. Take your time and make your choice carefully.

As in Wason & Shapiro's experiment, the rule was in the universal form; e.g. 'Every time I go to Miami I travel by car'. The rule was presented in large type on a 10·3 × 15·4 cm note card. A particular subject received one of the four possible rules formed by permuting the two city names (Miami and Atlanta) and the two modes of transportation (plane and car). Each variation of the rule was presented four times in each group to control for any bias about the relationship between a city and a mode of transportation that subjects might entertain. Each of the four stimulus cards had a separate day of the week appearing in smaller type on the top of the card. The order of the four cards in the presentation array was randomized for each subject.

In the abstract condition, subjects were given a 16-card familiarization deck. Each card in this deck had a letter of the alphabet on one side and a single digit on the other side. The following instructions were read, 'Here are some cards, please have a look through them. On each card there is a letter on one side and a number on the other side.'

When the familiarization deck was inspected and returned, four cards labelled, respectively, D, K, 3, and 7, and a card with a rule specifying a relationship between a letter and a number were presented. (As in the thematic problem, these cards were identical to four cards in the familiarization deck.) The following directions were read.

> I will now present you with four cards; the test you are about to do concerns only these four cards. These cards show a letter on one side and a number on the other side. I am going to make a claim about the connection between letters and numbers and your task is to say which of the cards would need to be turned over to decide whether that claim is true or false. You may choose any or all of the cards. Take your time and make your choice carefully.

As for the thematic problem, the proper control measures, such as using all four possible variations of the rule, were employed. Subjects in the abstract-problem-first group were treated identically to the thematic-problem-first subjects except for the order of the problems.

Results and discussion

The frequencies of the various selection combinations for each type of problem and for each trial are given in Table 1. There were only 2/16 correct solutions for thematic problems when presented first and only 1/16 correct solutions when presented second. There were no correct solutions for the abstract problems on either trial. No significant difference between correct response rates for thematic and abstract problems for either trial was found (Fisher's exact probability tests, one-tailed). In these and subsequent statistical tests reported in the paper, $P = 0{\cdot}05$ was the level of significance employed.

In accordance with Evans (1977) and Pollard (1981), the presentation of the data in terms of selection frequencies for each card and each trial is given in Table 2. There are no significant differences for p, $\bar{p}$, q, or $\bar{q}$ for either type of problem or trial number (Fisher's Exact Probability tests, two-tailed). Thus, for our data there were no effects of thematic

Table 1. Frequencies of various selection combinations as a function of type of problem and trial number ($n = 32$)

Selection combination	Type of problem			
	Thematic		Abstract	
	Trial 1	Trial 2	Trial 1	Trial 2
p	7	7	7	5
p, q	5	5	8	10
$p, q, \bar{q}$	1	2	0	0
$p, \bar{q}$	2	1	0	0
Other	1	1	1	1

Table 2. Selection frequencies of individual cards as a function of type of problem and trial number ($n = 32$)

Individual card	Type of problem			
	Thematic		Abstract	
	Trial 1	Trial 2	Trial 1	Trial 2
p	15	16	16	16
$\bar{p}$	0	1	1	1
q	6	8	9	11
$\bar{q}$	4	3	1	1

content for number of correct solutions or selection frequencies for individual cards. These data are in agreement with those of Manktelow & Evans (1979) and Yachanin (1980). Given the present results, there are now three recent replication studies (Manktelow & Evans, 1979; Yachanin, 1980; and the present experiment) that have found no effect for thematic materials of the Wason & Shapiro variety, two studies (Gilhooly & Falconer, 1974; and the more recent, Pollard, 1981) that found a statistically significant but weak effect, and two older studies (Bracewell & Hidi, 1974; Van Duyne, 1974) that found substantial effects in some but not all of their thematic-content conditions. In summary, if such material truly results in improvement in performance in the selection task, it does not do so consistently.

Experiment 2

One point should be made about the methodology of the Johnson-Laird *et al.* (1972) experiment. When describing the Johnson-Laird *et al.* experiment, many secondary sources (e.g. Wason & Johnson-Laird, 1972; Johnson-Laird & Wason, 1977; Wason, 1977; Anderson, 1980) only mention four of the five different envelopes that were actually used in each condition of the study (see Johnson-Laird *et al.*, 1972, p. 397). If the Johnson-Laird *et al.* experiment was conducted as described in their method section, then it contained a serious confounding. For their thematic rules [Numbers (1) and (2)], the fifth envelope (the unstamped, addressed, face-uppermost envelope) was clearly a $\bar{q}$-instance. However, *unless* instructions that letters (or numbers) could only appear on one side of a 'letter' (envelope)

were given, the nature of the fifth envelope (the face-uppermost, blank envelope) in the abstract conditions is ambiguous. Rules (3) and (4) do not tell the subject whether letters (or numbers) can *only* appear on one side of the 'letter'. Thus, the fifth envelope could either be an instance of $\bar{p}$ or of $\bar{q}$. It might be argued that the subject could *infer* that letters always occur on the front of the 'letter' and that numbers only occur on the back since this is the case for the other four envelopes. However, this would not be a valid inference because the stated rules say nothing about this.

Regardless, given this inference or the necessary instructions, the fifth envelope would be an instance of $\bar{p}$ for rules (3) and (4). Thus, the arrays used in the abstract conditions were different from those in the thematic conditions. If the inference were made, or the proper instructions given, $p\ \bar{p}\ \bar{p}\ q\ \bar{q}$ arrays were used in the abstract conditions while $p\ \bar{p}\ q\ \bar{q}\ \bar{q}$ arrays were used in the thematic conditions. Two $\bar{q}$-instances in the thematic arrays highlight a selection that is critical for the correct solution. This difference in arrays may have been very important, especially given Roth's (1979) recent finding of facilitation for reduced arrays with two $\bar{q}$-instances. Without the inference or instructions, the abstract conditions would be even more confusing than usual.

We dealt with this problem in the following way. We removed the ambiguity about the blank envelope in the abstract condition by means of a pre-task familiarization deck and instructions informing the subject as to what could appear on the front and the back of the envelopes. In addition, we only put letters on the backs of the envelopes and only numbers on the fronts. Thus, the fifth envelope would be an instance of $\bar{q}$ (as in the thematic conditions).

Because we wanted to conduct as precise a replication as possible, except for dealing with this ambiguity in the way described, the only major change in procedure that we made in the experiment was the use of American and Mexican stamps and addresses in place of English and Italian stamps and addresses. As in Expt 1, this change was made in order to make the problems more appropriate for our subject population.

Method

Subjects. Twenty-four undergraduates at the University of Florida participated as part of the requirements for a course in introductory psychology. None of these subjects had participated in Expt 1.

Design. As in Johnson-Laird *et al.* (1972), each subject acted as his/her own control and performed four versions of the selection task. The four versions were formulated by combining two types of problem content (thematic vs. abstract) with two types of syntactic construction (if–then and only–if) for the implication rule. The four rules used were: (1) If a letter is sealed, then it has a 15-cent stamp on it (thematic, if–then); (2) A letter is sealed only if it has a 3-peso stamp on it (thematic, only–if); (3) If a letter has an A on one side, then it has a 3 on the other side (abstract, if–then); and (4) A letter has a D on one side only if it has a 5 on the other side (abstract, only–if). Thus, the lexical material was varied across conditions. In addition, the order of presentation of the conditions was counterbalanced by the use of a Latin square design.

Materials. Five ordinary envelopes (9 × 13 cm) were used in each of the four conditions. For Rule 1, they had American addresses, and stamps were 10 or 15 cents. For Rule 2, they had Mexican addresses and the stamps were 2 or 3 pesos. Addresses and stamps were always on the front of the envelopes. For Rule 3, the five envelopes consisted of the following items: 'A' on the back of an envelope, 'B' on the back of an envelope, '3' on the front of an envelope, '2' on the front of an envelope, and a blank envelope, face uppermost. An equivalent set of stimuli was prepared for Rule 4, except the letters 'C' and 'D' and the numbers '4' and '5' were used.

Two familiarization decks of eight envelopes each were used. The deck for the problems with thematic content consisted of the following eight envelopes—two with American addresses and stamps (one sealed, one not sealed), two with foreign (not Mexican) addresses and stamps (one

sealed, one not), and four envelopes without addresses and stamps (two sealed, two not). The American stamps used in this familiarization stage were of different denominations from those used in the experimental tasks.

The following instructions were read as the deck was inspected.

> Notice that some of the envelopes are sealed and that some of them are not sealed. If a letter is laying face up on the table you cannot actually see whether or not it is sealed. Also notice that on the front of the envelopes there is not a stamp or there is a stamp of one value or another.

The familiarization deck for the abstract conditions consisted of four envelopes with numbers on the front and letters on the back, two envelopes with numbers on the front and nothing on the back, and two envelopes with nothing on the front and letters on the back. The subjects were read the following instructions as they inspected the deck.

> Notice that on the back of these envelopes either a letter is printed or the letter has been left off. Also notice that on the front there is either a number or the number has been left off. Remember that whenever there is a letter, it is on the back and whenever there is a number, it is on the front.

Procedure. Each subject was run individually. The appropriate familiarization deck and instructions were given to the subject before the presentation of the first problem for each of the two types of content. When subjects encountered a thematic-content problem, they were given the following instructions along with the rule.

> On this task imagine you are a post office worker sorting letters. It is your job to ensure that they conform to the following rule. Select those envelopes that you definitely need to turn over to find out whether or not they violate the rule. A letter that is face up may be sealed or it may not be.

These instructions were read to the subjects, and they were asked to read them again to themselves and to respond when they were ready.

On the problems with abstract content, the following instructions were on the card with the rule.

> On this task you must examine some envelopes. Here is a rule about the envelopes. Select those envelopes that you definitely need to turn over to find out whether or not they violate the rule.

The order of the various envelopes in the presentation array for each rule was randomized across subjects. After completion of the four tasks, the subjects were asked whether the problems were similar and, if so, in what way.

Results and discussion

Table 3 shows the number of subjects making zero, one, or two correct answers as a function of type of problem content. The results of Johnson-Laird *et al.* (1972) are included in the table for comparison purposes. Of particular interest is our complete failure to replicate the good performance for the thematic rules; subjects were correct on only 5 of 48 thematic problems.* This is to be compared to 39 of 48 correct for thematic content in the Johnson-Laird *et al.* experiment. No significant difference in correct response rates for thematic and abstract problems was observed on any of the four trials (Fisher's exact probability tests, one-tailed).

Table 4 presents the frequencies of various selection combinations as a function of problem type and syntactic construction. These frequencies show two things. First, as in Johnson-Laird *et al.* (1972), we observed little effect of the verbal formulation of the rule upon performance. Second, the particular selection combinations in our experiment were much more variable than usually observed in selection-task studies. The 'other' category represents seven or eight different selection combinations for each of the four rules. Similar variability in responding has also been observed by Yachanin (1980). In addition, no

* This complete failure to replicate the results of Johnson-Laird *et al.* (1972) was also observed in a pilot study we conducted. Twenty-four additional subjects produced only 2 of 48 correct solutions for thematic-content problems like those used in Expt 2. This pilot study differed from Expt 2 only in that no familiarization instructions and decks were employed. Thus, this pilot experiment was even more like the original Johnson-Laird *et al.* study, and no thematic-materials effect was observed.

Table 3. The numbers of subjects making a correct selection on both trials, one trial, and neither trial as a function of type of problem in Expt 2 and in Johnson-Laird *et al.* (1972) ($n = 24$ in both studies)

	Type of problem	Experiment 2	Johnson-Laird *et al.*
Both correct	Thematic	1	17
	Abstract	0	0
One correct	Thematic	3	5
	Abstract	1	7
Neither correct	Thematic	20	2
	Abstract	23	17

Note. A correct selection is defined as p, $\bar{q}_1$, and $\bar{q}_2$.

Table 4. Frequencies of various selection combinations as a function of problem type and syntactic construction ($n = 24$)

Type of problem	Syntactic construction	p	p,q	$p,q,\bar{q}_2$	$p,q,\bar{q}_1,\bar{q}_2$	$p,\bar{q}_1,\bar{q}_2$	Other
Thematic	If–then	5	4	1	0	2	12
	Only–if	3	3	1	1	3	13
Abstract	If–then	4	5	1	0	1	13
	Only–if	4	7	1	1	0	11

Note. No subject selected $p,q,\bar{q}_1$ so it is not included in the above table. $\bar{q}_1$ refers to the understamped envelope and $\bar{q}_2$ to the envelope with no stamp in the thematic problems. In the abstract problems, $\bar{q}_2$ is the blank envelope.

significant differences between the four rules were found for individual card choices on any of the four trials (Fisher's exact probability tests, two-tailed).

There was no reason to check for positive transfer since there were no significant differences between the four rules with respect to correct response rates and individual card choices. Likewise, during the post-task interviewing, none of the subjects reported noticing any similarity in the logical structure of the four problems.

Experiment 3

Manktelow & Evans (1979) argue that the impressive results of Johnson-Laird *et al.* (1972) may have been due to long-term memory cues – the falsifying instance (a sealed but under-stamped letter) would be available immediately from a subject's past experience. We also believe this to be a major factor in accounting for Johnson-Laird *et al.*'s data and for our failure to replicate these results. A conditional relationship between sealing a letter and the amount of postage on a letter is not a part of our subjects' past experience as it clearly was for the British subjects in the early 1970s. No postal regulation for letters concerning amount of postage and the sealing of envelopes exists here in the United States. Thus, such thematic material would not cue the falsifying instance in our subjects' memories as it could have in the memories of the subjects in the Johnson-Laird *et al.* study.

It should be noted that Johnson-Laird *et al.* observed no difference between thematic rules employing English stamps and those using Italian stamps. However, it is not difficult

to assume subjects knowing of the British regulation concerning 4d and 5d stamps would use this knowledge in dealing with Italian 40 and 50 lire stamps, especially since the largest denomination in both cases was the correct postage for sealed letters at that time and the amounts for both types of stamps had analogous aspects (4 and 5; 40 and 50).

If the memory-cueing hypothesis is correct, we should be able to facilitate performance in the selection task by using an implication rule that is a part of our subjects' experience. Our subjects are university undergraduate students from approximately 18 years to 22 years of age. An implication rule (actually a law in the state of Florida) with which most of these students are very familiar concerns the legal age for drinking alcoholic beverages – a person must be 19 years of age to legally drink such beverages. With such a rule, the falsifying instance (a person who is drinking an alcoholic beverage and who is under 19 years of age) is clearly part of our undergraduates' experience (evidence for this assumption will be presented later in the general discussion section). Improved performance for such a rule would support Manktelow & Evans's memory-cuing hypothesis. This drinking-age rule was examined in Expt 3.

Method

Subjects. Forty undergraduates at the University of Florida participated as part of the requirements of a course in introductory psychology. None of these subjects had participated in any of the earlier experiments.

Design. The design was identical to that employed in Expt 1. Each subject served as his/her own control. Twenty of the subjects were assigned to the group that received the thematic problem first while the other 20 subjects were assigned to the group that received the abstract problem first. Subjects were run individually.

Materials and procedure. One thematic problem and one abstract problem were used. The thematic problem (hereafter sometimes referred to as the 'drinking-age problem') was presented to the subject on a 10·3 × 15·4 cm card in the following manner.

> On this task imagine that you are a police officer on duty. It is your job to ensure that people conform to certain rules. The cards in front of you have information about four people sitting at a table. On one side of a card is a person's age and on the other side of the card is what the person is drinking. Here is a rule: IF A PERSON IS DRINKING BEER, THEN THE PERSON MUST BE OVER 19 YEARS OF AGE. Select the card or cards that you definitely need to turn over to determine whether or not the people are violating the rule.

Four other note cards labelled, respectively, DRINKING A BEER, DRINKING A COKE, 16 YEARS OF AGE, and 22 YEARS OF AGE were presented along with the instructions and rule card. The abstract problem was presented on a card as follows.

> The cards before you have information on both sides. On one side of a card is a letter, and on the other side is a number. Here is a rule about the cards: IF A CARD HAS AN 'A' ON ONE SIDE, THEN IT HAS A '3' ON THE OTHER SIDE. Select the card or cards that you definitely need to turn over to determine whether or not they are violating the rule.

The response cards were labelled, respectively, A, B, 2, and 3.

As in the earlier experiments, the order of the response materials was randomized in the presentation array to control for the possible effects of having a particular response card appear in a particular position. After the rule card and the four response cards were presented, the rule card was read to the subject who was then instructed to read the rule card silently and ask questions if needed.

Results and discussion

The frequencies of the different selection combinations are given in Table 5 for both types of problems for each trial. The results are very different from those found in our earlier experiments. Seventy-three per cent of the subjects made the correct selection for the drinking-age problem while no one did so for the abstract problem. On Trial 1, 14 of 20, and on Trial 2, 15 of 20 selections were correct for the drinking-age problem while no

Table 5. Frequencies of various selection combinations as a function of type of problem and trial number ($n = 40$)

Selection combination	Type of problem			
	Thematic		Abstract	
	Trial 1	Trial 2	Trial 1	Trial 2
p	5	3	3	0
p,q	0	0	8	13
$p,q,\bar{q}$	1	0	1	0
$p,\bar{q}$	14	15	0	0
Other	0	2	8	7

Table 6. Selection frequencies of individual cards as a function of type of problem and trial number ($n = 40$)

Individual card	Type of problem			
	Thematic		Abstract	
	Trial 1	Trial 2	Trial 1	Trial 2
p	20	18	14	13
$\bar{p}$	0	1	6	6
q	1	0	14	15
$\bar{q}$	15	17	8	6

correct solutions were obtained for the abstract-content problem. The difference in the proportion correct between thematic and abstract problems was significant for both trials ($\chi^2 = 18{\cdot}6$ and $\chi^2 = 20{\cdot}9$, respectively, $P < 0{\cdot}01$). There was not a significant difference in proportion correct on the thematic problem between the group receiving the problem first and the group receiving it after responding to the abstract problem.

The selection frequencies of individual cards as a function of type of problem are given in Table 6. The pattern of choices for the two problems are clearly different. Relative to responses on the abstract problem, the effect of the thematic content on Trial 1 was to significantly increase the proportion of subjects selecting the p card ($\chi^2 = 4{\cdot}9$, $P < 0{\cdot}05$), decrease the selection of the $\bar{p}$ card ($\chi^2 = 4{\cdot}9$, $P < 0{\cdot}05$), and decrease the selection of the q card ($\chi^2 = 15{\cdot}4$, $P < 0{\cdot}01$). The increase in the number of subjects selecting the $\bar{q}$ card approached significance ($\chi^2 = 3{\cdot}7$, $P < 0{\cdot}10$). On Trial 2 there was only a significantly higher proportion of $\bar{q}$ selections and a lower proportion of q selections on thematic problems ($\chi^2 = 10{\cdot}2$ and $\chi^2 = 20{\cdot}9$, respectively, $P < 0{\cdot}01$).

Because performance on the thematic problem was so good, we checked for possible transfer effects to the abstract problem when it was presented second. However, as can be clearly seen in Tables 5 and 6, no positive transfer effects were observed. This complete lack of transfer was also observed by Johnson-Laird *et al.* (1972).

These results are in total agreement with the memory-cueing hypothesis. However, even though there were 40 subjects in this experiment, another group of 18 subjects was given

the drinking-age problem to see if our finding of facilitation could be replicated. This seemed necessary given the inconsistency in successfully replicating past studies that have found good performance on thematic problems. Fourteen of these 18 subjects (74 per cent) made the correct selection. Thus, the improved performance for the drinking-age problem is clearly a reliable result for our subject population.

General discussion

The present series of experiments examined the performance of American undergraduates on abstract and thematic variations of Wason's four-card selection task. Including the pilot study for Expt 2 and the replication of Expt 3, we calculated some overall correct response rates for these variations. The correct response rate for the standard abstract problem was less than 1 per cent. For the thematic problems concerning cities and modes of transportation, the correct response rate was 9 per cent while for the thematic problems involving postal materials, it was 7 per cent. A dramatic increase in correct response rate, however, was observed for the thematic problems concerning the drinking-age rule; the correct response rate for such problems was 74 per cent. In addition, this high level of performance for the drinking-age problem did not bring about any positive transfer on the standard abstract problem.

These results are consistent with a memory-cueing hypothesis that proposes performance on the selection task is significantly facilitated when the presentation of the task allows the subject to recall past experience with the content of the problem, the relationship expressed, and a counter-example to the rule governing the relationship. For a highly reliable facilitation effect, it may also be necessary to initiate the search for this counter-example by presenting the problem in a manner that cues the subject to use a 'detective set' strategy in searching for a violator of the rule (such strategies have been discussed by Lunzer, 1975; Van Duyne, 1974, 1976; and others).

To obtain further evidence in support of such an explanation and to substantiate our assumption that the drinking-age rule was clearly in our subjects' experience, we administered a questionnaire to a group of 33 additional subjects from the same subject population that served in our earlier experiments. This questionnaire allowed us to estimate the experience our subjects had with the thematic materials and relationships used in our earlier experiments. Briefly, the questionnaire asked the subject whether regulations concerning drinking beer and being of a certain age and sealing a letter and having a certain amount of postage on it existed and, if so, to write these regulations. These questions were followed by a series of multiple choice items designed to tap the familiarity in our subjects' experience of counter-examples to such regulations.

While there is no regulation in the United States requiring different postage for sealed and unsealed letters, when subjects were asked to write such a regulation if they thought one existed, 12 per cent wrote a rule consistent with the statement of the rule used in the thematic problems in Expt 2. One subject even reported being able to recall a specific instance of violating the regulation! In marked contrast to these results, 88 per cent of the subjects wrote a rule consistent with the drinking-age rule used in Expt 3. Seventy-six per cent of the subjects also reported that they had personally violated the rule more than once, and 97 per cent of the subjects could remember specific instances of someone other than themselves violating the drinking-age rule. Thus, the assumption that our subjects were familiar with the drinking-age rule and counter-examples to it was supported by these questionnaire data.

The memory-cueing hypothesis can account for many of the previous findings on the thematic-materials effect as well as the results of the present series of experiments. The very good performance reported by Johnson-Laird *et al.* can be explained by their subjects'

personal knowledge of the postal rule and its counter-examples.* They were certainly given the detective-set strategy when instructed to imagine they were postal workers responsible for ensuring letters conformed to a rule. The American subjects, on the other hand, have no real knowledge of such a postal regulation and its counter-examples; and thus, in Expt 2, they were unable to perform well on this type of thematic material even though a detective-set strategy was provided by the instructions. With the drinking-age problem, American subjects were given a problem that cued both knowledge of a rule and its counter-examples and the detective-set strategy. Thus, in Expt 3, there was a high percentage of correct selections for such problems.

Similarly, Van Duyne's unpublished data (cited in Van Duyne, 1976) on non-arbitrary postal rules can be explained as the result of memory-cueing. His data for arbitrary rules, however, require an additional assumption of the occurrence of transfer, either through direct experimental experience with the related, non-arbitrary postal rule or indirectly through reasoning by analogy from the related, non-arbitrary rule in memory. This latter process would be like that proposed earlier for the two thematic rules in Johnson-Laird *et al.* (1972).

We are currently examining this assumption about transfer with our drinking-age rule and an arbitrary, related rule (i.e. 'If a person is wearing blue, then the person must be over 19 years of age'). Our data indicate a correct response rate of about 75 per cent for a problem using this arbitrary rule and following a problem on the related, non-arbitrary rule. The correct response rate on the arbitrary-rule problem when given first is only about 25 per cent. Thus, these data agree with our assumption about transfer and support a memory-cueing explanation of Van Duyne's data for an arbitrary, thematic postal rule. It should also be noted that our preliminary data reveal no facilitative transfer effects for standard abstract problems from non-arbitrary, thematic problems. This finding agrees with the results of Expt 3 and those of Johnson-Laird *et al.* (1972).

Further support for the memory-cueing hypothesis was presented by Van Duyne (1976). He briefly reports the results of two pilot studies in which subjects performed selection tasks using 'non-arbitrary and obviously commonplace sentences' (e.g. 'If it is raining then the streets are wet'). Such rules may have been interpreted as necessarily true, thereby preventing subjects from looking for falsifying instances (counter-examples). If this were the case, performance would not have been good, and it was not. Van Duyne reports that it was 'surprisingly low'. Thus, the memory-cueing hypothesis would seem a plausible explanation for these data.

The inconsistency of the results for the other types of thematic material [cities (or locations) and modes of transportation, foods and beverages, and students' universities and major fields of study] makes them very difficult to explain. Sometimes a fairly substantial thematic effect is found, and at other times, it is completely non-existent. Such contradictory data are probably the result of a very complex interaction between subject and task factors. The inconsistency in results could be brought about by the unavailability of counter-examples in the subjects' experience, a particular problem presentation failing to cue the subject to search for counter-examples (as hypothesized for Van Duyne's pilot studies), the difficulty of finding counter-examples if available, or some combination of these factors.

Following this line of explanation, one can speculate about reasons for the

* This explanation of Johnson-Laird *et al.*'s results is remarkably consistent with some recent findings of Evelyn Golding, presented at the BPS Annual Conference, The University of Surrey, April 1981. She replicated Johnson-Laird *et al.*'s results using older British subjects who had experienced the relationship of a higher-priced stamp on a sealed envelope, but failed to replicate their results using younger British subjects who had no experience with the older postal system. Thus, in agreement with the memory-cueing hypothesis, knowledge of the older British postal system would appear to be crucial in producing Johnson-Laird *et al.*'s results.

inconsistencies that have been observed. For example, Stephen Newstead has suggested to us the following possibility. In the Wason & Shapiro study the University of London students might well be familiar with the disconfirming instances (going to Leeds and Manchester by car or train) through their past experience. However, the Plymouth Polytechnic subjects in the Manktelow & Evans (1979) and Pollard (1981) experiments might be less familiar with such journeys since they would be farther away from the cities involved and probably less likely to have actually made the journeys. Such speculation could explain why the Plymouth subjects would show a much reduced effect in the Pollard study or no effect in the Manktelow & Evans study. While such speculation is an interesting endeavour, more direct experimental evidence is obviously needed to tease out the complex interaction that must be at the base of this set of inconsistent findings.*

In summary, the evidence for the thematic-materials effect that cannot be directly attributed to memory-cueing is weak and inconsistent. When the data indicate a strong effect, it can almost invariably be attributed to memory cueing and not facilitation of logical reasoning. In referring to some of the studies on the thematic-materials effect, Wason (1977, p. 132) comments, 'it is a little depressing too because it suggests that the critical appreciation of falsification is aroused, not by invoking a calculus, but by assimilation to more mundane experience'. It is this 'mundane experience' that we feel is responsible for the thematic-materials effect.

Acknowledgements

The authors would like to thank Carlos and Jane Galina for providing us with the Mexican stamps and addresses used in Expt 2 and Steve Newstead for helpful comments on an earlier version of this paper.

References

Anderson, J. R. (1980). *Cognitive Psychology and Its Implications*. San Francisco: Freeman.

Bracewell, R. J. & Hidi, S. E. (1974). The solution of an inferential problem as a function of stimulus materials. *Quarterly Journal of Experimental Psychology*, **26**, 480–488.

Brown, C., Keats, J. A., Keats, D. M. & Seggie, I. (1980). Reasoning about implication: A comparison of Malaysian and Australian subjects. *Journal of Cross-Cultural Psychology*, **11**, 395–410.

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

Evans, J. St B. T. (1978). The psychology of deductive reasoning: Logic. In A. Burton & J. Radford (eds), *Thinking in Perspective*. London: Methuen.

Gilhooly, K. J. & Falconer, W. A. (1974). Concrete and abstract terms and relations in testing a rule. *Quarterly Journal of Experimental Psychology*, **26**, 355–359.

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. & Wason, P. C. (1977). A theoretical analysis of insight into a reasoning task. In P. N. Johnson-Laird & P. C. Wason (eds), *Thinking: Readings in Cognitive Science*. Cambridge: Cambridge University Press.

Lunzer, E. A. (1975). The development of advanced reasoning abilities. *Italian Journal of Psychology*, **2**, 369–390.

Lunzer, E. A., Harrison, C. & Davey, M. (1972). The four-card problem and the generality of formal reasoning. *Quarterly Journal of Experimental Psychology*, **24**, 326–339.

Manktelow, K. I. & Evans, J. St B. T. (1979). Facilitation of reasoning by realism: Effect or non-effect? *British Journal of Psychology*, **70**, 477–488.

Pollard, P. (1981). The effect of thematic content on the 'Wason selection task'. *Current Psychological Research*, **1**, 21–29.

Reich, S. S. & Ruth, P. (1982). Wason's selection task: Verification, falsification and matching. *British Journal of Psychology*, **73**, 395–405.

* Since the submission of this manuscript, two additional failures to find a thematic-materials effect have been brought to our attention – Brown *et al.* (1980) and Reich & Ruth (1982). Brown *et al.* used towns-and-transport content as in Wason & Shapiro (1971). Reich & Ruth used both food-and-drinks content as in Manktelow & Evans (1979) and coherent scenarios similar to Wason & Shapiro's towns-and-transport material. The results of these two studies in addition to those of Manktelow & Evans (1979), Yachanin (1980), and the present study make the Wason & Shapiro findings even less robust and thus less interesting as a topic for further investigation.

Roth, E. M. (1979). Facilitating insight in a reasoning task. *British Journal of Psychology*, **70**, 265–271.
Van Duyne, P. C. (1974). Realism and linguistic complexity in reasoning. *British Journal of Psychology*, **65**, 59–67.
Van Duyne, 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. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, **20**, 273–281.
Wason, P. C. (1977). The theory of formal operations – A critique. In B. A. Geber (ed.), *Piaget and Knowing: Studies in Genetic Epistemology*. London: Routledge & Kegan Paul.
Wason, P. C. & Johnson-Laird, P. N. (1972). *Psychology of Reasoning: Structure and Content*. Cambridge, MA: Harvard University Press.
Wason, P. C. & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, **23**, 63–71.
Yachanin, S. A. (1980). Differential effects of thematic materials and order of rule presentation on a reasoning task. Unpublished Masters thesis, Bowling Green State University, Bowling Green, Ohio, USA.

Received 6 February 1981; revised version received 20 July 1981

Requests for reprints should be addressed to Richard A. Griggs, Department of Psychology, University of Florida, Gainesville, Florida 32611, USA.
James R. Cox is also at the above address.