CHAPTER 2

A review of the literature on the "elusive" content effect on the Wason selection task

When content effects are found, the hypothesis that they were produced by content-dependent cognitive processes should be entertained. The extensive literature on the Wason selection task is replete with reports of content effects. If content-dependent inference procedures exist, this literature is a promising place to look for them.

Attempts to predict and explain content effects on the Wason selection task in terms of "differences in subjects' experience" with the different content domains tested have created a hornet's nest of apparently contradictory results. The unpredictability and unreplicability of the content effect on the Wason selection task is so pronounced that Peter Wason has called it a "crisis" (personal communication) and Griggs and Cox (1982) have dubbed the effect "elusive."

Because the explanation of this elusivity is the subject of my thesis, this chapter will explore these results in some detail, one content area at a time. If domain specific reasoning processes are involved, then the data should resolve into patterns when it is categorized by content domain per se, but not when it is categorized by factors correlated with content, like "familiarity" or "realism." Five major content areas have been explored in the literature: transportation, food, school, and "social contracts."

The discussion of published explanations attempting to
2.1 The Transportation Problem

The "Transportation Problem", developed by Wason & Shapiro (1971), has been used in more experiments testing for an effect of thematic content in the Wason selection task than any other thematic rule. It is a conditional rule linking a place to a means of transportation, for example, "If I go to Boston, then I take the subway." Researchers always use places and means of transportation that are local for and familiar to their subject population. There are nine experiments comparing performance on the transportation problem to performance on an abstract problem. Two found substantial content effects, two found weak content effects, and five found no content effects at all.

Wason & Shapiro, 1971

The first demonstration of a content effect was by Wason & Shapiro (1971). They gave 16 subjects a selection task using the rule "Every time I go to Manchester I travel by car" (thematic group), and 16 subjects a selection task using the rule "Every card which has a vowel on one side has an even number on the other side" (abstract group).* Destinations and means of transportation were rotated in the thematic group to avoid the possibility of an effect due to preconceptions about the relation between particular destinations and means of transport. Sixty-two

* Only Wason & Shapiro's abstract problem used the vowel-even number rule. Abstract problems in the other studies linked specific letters and numbers, e.g., "If there is an 'A' on one side of a card, then there is a '3' on the other side."
percent of the thematic group gave the logically correct, falsifying answer, 'P & not-Q', whereas only 12% of the abstract group gave this answer (\(\phi = .52\)).

Bracewell & Hidi, 1974

In 1974, Bracewell & Hidi and Gilhooly & Falconer tried to replicate Wason & Shapiro, 1971. Their experiments were designed to tease apart the relative contribution to success on the selection task of concrete terms versus concrete relations. Here I will only discuss the conditions that are directly comparable to Wason & Shapiro's thematic and abstract groups, because establishing the existence of a content effect is theoretically prior to asking what causes it.*

Noting that the most common selection task error is to incorrectly select the card corresponding to the Q term, Bracewell & Hidi wondered if subjects "spend more time analysing the first set of terms to the detriment of the second set." They tested this by framing their thematic and abstract problems in two different linguistic formats: "Every time P, Q" and "Q every time P" (e.g., "Every time I go to Ottawa I travel by car" and "I travel by car every time I go to Ottawa."). The logical structure of these two problems is identical, however the Q term comes first in the "Q every time P" format. Their results are pictured in Table 2.1.

The "Every time..." linguistic format (also used by Wason & Shapiro) successfully replicated Wason & Shapiro's findings: 9 out of 12 subjects (75%) answered 'P & not-Q' in response to the thematic rule versus 1 out of 12 (8%) for the abstract rule

* The relative contribution question will be considered in Chapter 3.
Table 2.1 Results of Bracewell & Hidi, 1974

<table>
<thead>
<tr>
<th></th>
<th>thematic</th>
<th>abstract</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Every time P, Q:</td>
<td>9</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Q every time P:</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Totals:</td>
<td>11</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Number of subjects choosing 'P & not-Q'; n=12 per cell.

(\( \phi = .68 \)). However, there was no thematic content effect for the "Q every time P" phrasing; 2 out of 12 subjects in the thematic group falsified (17%), compared to 1 out of 12 (8%) in the abstract group. There is no reason to believe that this second phrasing is an unnatural one*; in fact, this is the phrasing which Bracewell & Hidi had hoped would enhance logical performance by focusing attention on the Q term.

Thus, a simple change in linguistic format completely erased the content effect.

Interpretation of Bracewell & Hidi's data is further complicated by the fact that they explicitly told their subjects that the conditional is not "reversible." This instruction is unprecedented in selection task research; so are its apparent effects. The most common response to abstract problems is usually 'P & Q'. Yet only 1 out of the 24 subjects in Bracewell & Hidi's two abstract conditions gave this response, and no one gave it in either of the thematic conditions.

This instruction is so serious a confound that some

* I would guess that pragmatic factors determine which phrase would come first in ordinary conversation -- whether the speaker wished to indicate that the topic of the sentence is going to Manchester ("Every time I go to Manchester...") or traveling by car ("I travel by car...").
researchers are hesitant to count Bracewell & Hidi's "Every time" condition as a replication of Wason & Shapiro (Manktelow, 1979; Griggs & Cox, 1982; Griggs, 1983). I believe it may have introduced demand characteristics of the following kind.

When stumped by a problem, people sometimes ask "what's the trick?" -- it is a request for insight into the problem. However, I have never heard anyone stumped by a problem ask, "what are the tricks?" In other words, people usually assume that a thought problem has one "trick", not two. But solving the selection task involves two "tricks": according to Wason & Johnson-Laird (1972), subjects have not achieved "complete insight" unless they realize 1) that the Q card is irrelevant, and 2) that the not-Q card is relevant. When I was conducting pilot experiments, subjects who had finished the task frequently asked me if the "trick" was realizing that one should omit Q. And, in fact, the second most common response on the selection task is 'P' alone, omitting the Q card.

If you believe you have found the "trick", why look for a second one? Telling subjects that the conditional is "not reversible" may be giving away half the game. When the meaning of "not reversible" is clear, this instruction is equivalent to telling them to omit the Q card. The task of finding "the trick" remains.

This could explain why subjects found the "not-Q trick" in the "Every time P, Q" format, but not in the "Q every time P" format. When the logical operator that defines the conditional is at the beginning of the sentence, as in "Every time I go to Ottawa I travel by car", the meaning of "reverse" is
straightforward. Subjects have virtually been told to omit Q, so they continue to search for the problem's trick -- choosing not-Q and may eventually find it. But the meaning of "reverse" is far more ambiguous when the logical operator is in the center of the sentence. What is the "reverse" of "I travel by car every time I go to Ottawa"? Is it "Every time I go to Ottawa I travel by car" or is it "I go to Ottawa every time I travel by car"? Figuring this out may have been challenging enough to count as "the trick" for subjects solving problems in the 'Q every time P' format. Telling them the conditional was "not reversible" was enough of a clue to allow most of them to finally figure out that they were supposed to omit Q, but having realized that, they stopped their search -- they thought they had found the "trick."

If we leave aside this methodological objection, Bracewell & Hidi's experiments can be thought of as two separate attempts to replicate Wason & Shapiro: one a success, the other a failure.

Gilhooly & Falconer, 1974

The design of Gilhooly & Falconer was similar to Bracewell & Hidi, except they used only the "Every time P,Q" linguistic format, and many more subjects (n=50 per group).

Gilhooly & Falconer's thematic group did significantly better than their abstract group: 22% v. 6% chose 'P & not-Q'.* However, the success rate for the thematic condition was quite low: 22%, as compared to 62% for Wason & Shapiro and 75% for

*Gilhooly & Falconer were puzzled that the error responses which Johnson-Laird & Wason (1970) classify as "partial insight" ('P, Q, and not-Q') did not have the same distribution as the "complete insight" ('P & not-Q') responses. If one grants their assumption that these two scores express progressively greater degrees of insight into the logical structure of the problem, and therefore lumps them together, the content effect disappears (26% v. 18%)
Bracewell & Hidi's best group. The effect size, $\phi = .23$, for Gilhooly & Falconer is closer to the "effect" size of .13 for Bracewell & Hidi's no effect condition than it is to the phi of for Wason & Shapiro. It is not unheard of for 22% of a subject population to get the abstract problem correct; in a number of my experiments, more than 22% of subjects falsified on the abstract problem (see Chapter 6). Furthermore, because Gilhooly & Falconer's sample size is three times larger than either Wason & Shapiro's or Bracewell & Hidi's, one might expect their figures to be somewhat less subject to Type 1 errors.

Thus, if one counts any facilitation with thematic content, however small, as a "content effect", then Gilhooly & Falconer counts as a replication of Wason & Shapiro. However, if by "content effect" one means that a majority of subjects give a logically correct response with a thematic rule, then Gilhooly & Falconer have failed to replicate Wason & Shapiro. As will be discussed in the next chapter, the theoretical claim being made determines which definition is appropriate.

Pollard, 1981

In a very close replication of Wason & Shapiro's initial study, Pollard (1981) found a mild content effect: 4 out of 12 subjects (33%) in the thematic condition gave the logically correct answer, whereas none of the 12 subjects in the abstract condition gave this answer ($\phi = .45$).

It is worth noting that given the percentage difference between the two groups (33%), zero correct in the abstract condition is the only outcome that could yield a significant result. If the same percent difference is maintained, but just
one subject in the abstract condition had answered correctly (hence 1 out of 12 in the abstract group, compared to 5 out of 12 in the thematic group), the difference between the two conditions would be insignificant (p < .07, Fisher's Exact). Given Pollard's small sample sizes, such a precarious result should be interpreted with caution.

This, so far, has been the good (and lukewarm) news for the content effect with a transportation problem. Now, the bad news.

Manktelow & Evans, 1979

In 1979, Manktelow & Evans conducted an experiment (Experiment 5) that duplicated Wason & Shapiro (1971) in every respect except one: they used an "If-then" linguistic format instead of the "Every time" format used by Wason & Shapiro (1971), Bracewell & Hidi (1974), Gilhooly & Falconer (1974), and Pollard (1981). Performance for the thematic and abstract groups was identical.

Brown, Keats, Keats, & Seggie, 1980

Brown, Keats, Keats, & Seggie (1980) also tried to replicate Wason & Shapiro, 1971, using 24 Australian and 24 Malaysian university students. Like Wason & Shapiro, their transportation problem used an "Every time" linguistic format. Their abstract problem used shapes and letters: "Every card with a black triangle on one side has a Y on the other side." For Malaysian subjects, the selection task was translated into Bahasa Malaysia, their national language. For both problems, subjects were told that the variables were strictly binary (travel is only by car or airplane, trips are only to Singapore or Penang), thus reducing the array of possible combinations of values from an infinite set.
to four. Unlike other transportation problem experiments, Brown et al. did not rotate destinations and means of transportation. Half the Australians and half the Malaysians were given the transportation problem; the remaining subjects were given the abstract problem. Brown et al. found no enhancement of logical performance with thematic content. None of the Malaysian subjects answered either problem correctly, one Australian answered the abstract problem correctly, and two answered the thematic problem correctly (the between-cultures factor was not significant).

Yachanin & Tweney, 1982

Yachanin & Tweney (1982) looked in vain for a thematic content effect in a variety of different content areas. Evans & Lynch (1973) argued that performance on selection tasks with abstract rules is guided by a "matching bias": a tendency to choose cards that match values mentioned in the rule, regardless of their logical status (i.e., regardless of whether the propositions in the rule are affirmative or negative). This can only be tested by systematically negating components of a rule. Thus, given the rule "If not-A then not-3", subjects would choose the "A" and "3" cards because they are directly mentioned in the rule. By coincidence, choosing the "3" card is "logically" correct because it represents a false consequent (not-Q), and choosing the "A" card is logically incorrect because it represents a false antecedent (not-P). The matching bias is considered a non-logical response bias because it is blind to the logical structure of the problem.

Yachanin & Tweney argued that if thematic content truly
facilitates logical reasoning, then it should "protect" subjects from matching bias. Hence, they used four forms of each "If-then" rule: 1) affirmative antecedent and consequent (AA), 2) negative antecedent and consequent (NN), 3) affirmative antecedent and negative consequent (AN), 4) negative antecedent and affirmative consequent (NA). Subjects were tested on two of each of these rule forms (a total of eight problems per subject). A subject's eight problems were either all thematic or all abstract (n=40 per group). Yachanin & Tweney found no difference in performance between their thematic group and their abstract group for any of the rule forms (transportation: 13%, abstract: 11%). They did find evidence for both matching bias and a verification strategy in both groups.

One could argue that this result is uninteresting because there is evidence (reviewed in Wason & Johnson-Laird, 1972) that negatives are difficult to understand, hence subjects simply became confused in this experiment. There are two problems with this criticism. The first is theoretical: Many explanations of why there should be a thematic content effect are based on the idea that, by virtue of their familiarity, imageability, coherence, etc., thematic rules make confusing statements easier to understand. Thus, one would still expect a relative enhancement for thematic rules with simple scope negative components when compared to abstract rules with the same structure of negation, even if performance on these thematic rules is not as high as performance on affirmative rules. The second problem with this criticism is empirical: Even when one considers only affirmative (AA) rules, there is no difference in
card selections between abstract and thematic groups. One would have to argue that merely being exposed to a rule with a negative component is sufficient to totally confuse subjects when they then encounter an AA rule. Manktelow & Evans (1979) tested this possibility and found no evidence for it (see section 2.2 below).

Griggs & Cox, 1982

In 1982, Griggs & Cox tried to replicate Wason & Shapiro's result (Experiment 1). They used 32 subjects and gave each two problems: a transportation problem and an abstract problem. Half got one first, half the other. Like Wason & Shapiro, they used the "Every time" phrasing. Unlike Wason & Shapiro, they found no difference in performance between the two problems.

Transportation Problem Summary

The transportation problem elicited a substantial (greater than 50% falsification rate) content effect in two experiments (Wason & Shapiro, 1971; Bracewell & Hidi, 1974), a weak content effect in two experiments (Gilhooly & Falconer, 1974; Pollard, 1981), and no content effect at all in five experiments (Bracewell & Hidi, 1974; Manktelow & Evans, 1979; Brown et al., 1980; Yachanin & Tweney, 1982; Griggs & Cox, 1982).

2.2 The Food Problem

The "Food Problem" was developed by Manktelow & Evans (1979). It is a conditional rule about meals, linking something a person eats with something that person drinks, for example, "If I eat salad then I drink water." Performance on the food problem has been compared to performance on an abstract (or "low thematic") problem in six experiments, four by Manktelow & Evans

33
(1979), one by Yachanin & Tweney (1982), and one by Reich & Ruth (1982). No one has found an enhancement in logical performance with the food problem.

Manktelow & Evans, 1979

Manktelow & Evans conducted four experiments using food problems (Manktelow & Evans, 1979, Experiments 1-4). The protocol for their Experiment 1 was similar to that described in section 2.1 for Yachanin & Tweney (1982). They systematically varied which logical component was affirmative or negative. Each subject was given an AA, AN, NA, and NN problem. Like Yachanin & Tweney, Manktelow & Evans reasoned that if thematic content facilitates the use of deductive logic, then subjects will be less likely to fall victim to the matching bias when given a food problem than when given an abstract problem. Every subject was given a test booklet with written instructions and four problems: 24 subjects were given four food problems, 24 were given four abstract problems. The 48 subjects were tested at the same time, as a group.

Performance on the food problems was as low as performance on the abstract problems, and both groups showed evidence of the matching bias. This result holds even if one considers only affirmative (AA) rules.

Puzzled by this result, Manktelow & Evans systematically varied task factors that could have interfered with logical performance. Experiment 2 was identical to Experiment 1, except subjects were tested individually, alone in cubicles, rather than in a group. The results were the same as for Experiment 1. Testing subjects individually or in a group appears to have no
effect on their performance.

Next, Manktelow & Evans wondered if presenting so many rules, and rules with some negated components, was simply imposing too great a "cognitive load" on their subjects -- confusing them. So in Experiment 3, each of 32 subjects answered only one, affirmative (AA) selection task (half were given a food problem, half an abstract problem). Subjects were run in small, unsupervised groups. Again, there was no difference in card selections between the thematic and abstract groups.

Last, Manktelow & Evans wondered if previous enhancements in performance with thematic problems could have been due to the presence and participation of the experimenter. In most of the earlier studies, the experimenter had read the instructions aloud, allowed subjects to inspect a deck of sample cards from which the four cards for the selection task were randomly drawn, and requested and recorded subject responses. Manktelow & Evans' Experiment 4 was identical to their Experiment 3, except subjects were run as described above, with the experimenter controlling the whole sequence of events. Again, there was no difference in performance between the thematic and abstract groups.*

One last point: Using Manktelow & Evans' data on the frequency with which individual cards were chosen for thematic and abstract groups,** one can consider the hypothesis that thematic content reduces confusion, even if it does not facilitate

---

* Manktelow & Evans' Experiment 5, using the transportation problem (described in 2.1), was also conducted this way. The results were the same.

** They report the frequency with which individual cards were chosen, regardless of the combination in which they were chosen.
logic by increasing the probability that not-Q is chosen. Manktelow & Evans had no hypotheses regarding the direction of differences for the P, not-P, and Q cards, so for these they used two-tailed Fisher's Exact tests. But suppose thematic content reduces confusion, and choosing not-P or Q, or failing to choose P, is evidence of confusion. Then one would use a one-tailed test with the prediction that not-P and Q are chosen less frequently for thematic problems and P is chosen more frequently. Their data do not support this hypothesis. Using one-tailed tests, there are no cases of differential choosing of P cards. In Experiments 1, 2, 4, & 5, there are no differences in the choice of not-P or Q cards between thematic and abstract groups. In Experiment 3 the thematic condition elicited fewer not-P choices (p < .038, predicted direction), but more Q choices (p < .049, opposite of predicted direction). Thus, Manktelow & Evans' data provide no support for the hypothesis that thematic content decreases confusion about the conditional's meaning.

Yachanin & Tweney, 1982

Yachanin & Tweney's (1982) study included a condition identical to their transportation problem condition (described in section 2.1), except that the thematic group was tested on food problems (the abstract group used for comparison was the same as that for the transportation problem). They found no significant difference in responses between the thematic and abstract groups (food: 14%, abstract: 11%). This is true even if one considers only the affirmative (AA) problems.

Reich & Ruth, 1982

For lack of a better place, I include Reich & Ruth (1982),
in the food problem section. Their experiment differs from the others in that they looked at performance on "low thematic" versus "high thematic" problems, without using an abstract problem for comparison. Their "low thematic" problems were food problems. Their "high thematic" problems were embedded in a story context, for example: "Molly is employed at a seaside cafe. She has been instructed by her boss to serve tea or coffee only at certain times of the day. Visitors notice that: When it is early Molly serves tea. Are they correct?" The object was to create a coherent, "unitary", easy to visualize scenario linking the terms of the conditional.

Like Yachanin & Tweney (1982), Reich & Ruth gave each subject one affirmative and three negated forms (AN, NA, NN) of each rule. Twenty-four subjects were given low thematic rules, 24 were given high thematic rules. High thematicity did not significantly improve logical performance, whether one considers all four rule forms (low thematic: 17%; high thematic: 22%) or only the affirmative form (low thematic: 4%; high thematic: 17%). Moreover, performance on the "low thematic" food problems was in the same low range of values typically found for abstract problems.

**Food Problem Summary**

None of the six experiments testing food problems elicited an enhancement in logical performance with respect to either abstract or "low thematic" problems. Furthermore, although some researchers (e.g., Pollard, 1981) have claimed that food problems are not as "thematic" as transportation problems, no one has yet proposed a criterion for judging "thematicity", nor has any one
produced a reasoned argument to support the claim that food problems are less thematic. Indeed, considering that people eat and drink at several meals every day, one might think that, if anything, food themes should be more familiar to subjects than transportation themes.

2.3 The School Problem

The school problem, developed by Van Duyne (1974), is a conditional relating a person's major field of study to his or her school, for example, "If a student studies philosophy, then he goes to Harvard." There are two experiments studying this problem: one found better performance for school problems than for abstract problems, the other did not.

Van Duyne, 1974

Van Duyne compared performance on abstract and school problems for four logically equivalent linguistic formats:

Universal: "Every student who studies physics is at Oxford."

Standard Conditional: "If a student studies philosophy then he is at Cambridge."

Disjunctive: "A student doesn't study French, or he is at London."

Conjunctive: "It isn't the case that a student studies psychology and isn't at Glasgow."

All four linguistic formats are logically equivalent to the conditional "If a student studies field A, then he goes to university B" (you can convince yourself of this by consulting the truth table for the conditional in Chapter 1: e.g., the disjunctive sentence is equivalent to "If a student studies French, then he is at London"). For all four linguistic forms,
the correct answer is to choose the major field mentioned in the problem and the university not mentioned in the problem.

Van Duyne made no attempt to rotate combinations of fields and schools to avoid effects due to preconceptions; the four sentences above are the four school problems he used. Each of his 24 subjects answered the four problems above and four abstract problems that had the same linguistic formats. Half the subjects answered the four abstract problems first, the other half answered the four school problems first.

For the abstract problems, there were no significant differences in percentage correct among the four linguistic formats. Performance on the disjunctive and conjunctive forms of the school problem was as low as performance on the abstract problems. However, there was a difference in performance between the school and abstract problems when they were phrased as universals and as standard conditionals. For the universal phrasing, 58% of subjects gave the logically correct answer on the school problem as opposed to 8% on the abstract problem. For the standard conditional phrasing, 50% answered correctly on the school problem, but only 12% on the abstract problem.

As in Bracewell & Hidi's (1974) experiment (section 2.1, transportation problem), the content effect disappeared in some linguistic formats. However, the absence of a content effect for Van Duyne's disjunctive and conjunctive formats is less damning than it is for Bracewell & Hidi's "Q every time P" format, which is a minor, and pragmatically common, variation on the universal format. Van Duyne's disjunctive and conjunctive formats contain a number of complicating confounds. For example, in English, "A
or B" can mean "A or B but not both" or "A or B or both". Also, his disjunctive's first component is a negative, and as mentioned previously, there is evidence that people have difficulty interpreting negatives (Wason & Johnson-Laird, 1972; of course, one could still argue that thematic content should lessen the interpretational difficulties). Furthermore, Van Duyne's conjunctive not only has two negatives, but it has two negatives of different scope -- the first is meant to encompass the whole following statement, whereas the second applies only to the school. Last, "A student doesn't study French or he is at London" is a rather bizarre way of saying "If a student studies French then he is at London," and "It isn't the case that a student studies psychology and isn't at Glasgow" is a strange way of saying "If a student studies psychology then he is at Glasgow." Pragmatically, a negative is usually used to contradict a presupposition that is the topic of conversation -- it is not used to introduce a topic (Clark & Clark, 1982, p. 99). For these reasons, I do not find the lack of a content effect for these two linguistic formats very interesting. Too many other factors could be swamping the effect.

The universal and standard conditional formats did elicit content effects. However, I would like to offer two caveats in interpreting this result.

1) Any of the subjects taking this test know that in real life the rule expressed by the school problem is false. Universities are not segregated by major field. It is simply false that all psychology students go to Harvard -- some go to
Yale, Tufts, U. Mass., etc., and, I presume, every student knows this. Compounding this problem, Griggs (1983) has pointed out that in the U.K., Cambridge is renowned for its excellence in physics (the rule pairs physics with Oxford), and Oxford is renowned for its excellence in philosophy (the rule pairs philosophy with Cambridge).

This creates interpretational problems because there is evidence indicating that if a subject has personal beliefs about the veracity of the relation expressed by a logical problem, that subject's performance on the logical problem is guided, in part, by a desire or tendency to confirm those personal beliefs. In other words, when subjects believe a statement to be true they try to verify it, and when they believe it to be false they try to falsify it (Janis & Frick, 1943; Wason & Johnson-Laird, 1972, Ch. 7; Van Duyne, 1976; Pollard & Evans, 1981; see Pollard, 1982 for review). Given the reputations of the schools used, a "belief bias" would lead to falsification. Thus, one doesn't know if the effect is due to belief bias, or due to the effect of thematic content as such. In testing for content effects, one wants to neutralize any effects of belief bias, not exacerbate them by using rules that are both false and contrary to the subject's personal prejudices.

2) This problem has some (but not all) of the earmarks of a social contract problem, for which there does seem to be reliable evidence for a content effect. Briefly, in social contract problems the conditional rule expresses a contract in which a person is eligible for a benefit if, and only if, she pays a price or meets a requirement (fuller descriptions follow in
Given such a rule in a Wason selection task, a subject looking for a "cheater" -- an individual who has absconded with the benefit without having paid the price or met the requirement -- would choose the same cards as a subject seeking logical falsification.

Van Duyne's school rule was embedded in a "look for cheaters" context. Subjects were told that the cards were taken from a register of students who are eligible for a grant (the benefit), and that certain rules are supposed to apply to eligible students (the requirements that must be met to get the benefit). Rather than being asked to turn over the cards necessary to see "whether the rule is true or not" (a common wording), subjects were asked to turn over the cards necessary to see "whether they violate the rule or not" (emphasis mine). "Violate" was used for both the abstract and school problems, so it cannot, in itself, explain the difference. But my point is that this choice of words in conjunction with a context defining eligibility for a benefit suggests that one is looking for a violator, that is, a cheater on a social contract. Thus, it was not clear to me whether I should include Van Duyne's school problem in this section or the next one, on social contract problems. According to the formulation that will be presented in Chapter 5, a full fledged social contract should have the benefit stated in the antecedent: in this problem, the benefit is stated in the context, and the rule states a conditional requirement that earning the benefit is contingent upon. Thus, it is a hybrid between a full-fledged social contract and a straightforward relational rule with thematic content. This
makes it difficult to know whether the effect was due to the use of thematic content in general, or whether it was specific to the use of a social contract context.

Please note that if this context exercised a major effect it would mitigate the criticism expressed in the first caveat, because the rule would not be interpreted as a statement about all students, merely about those interested in earning the grant.

Yachanin & Tweney, 1982

Yachanin & Tweney's (1982) study included a condition identical to their transportation and food conditions (described in sections 2.1 and 2.2), except that the thematic group was tested on school problems (the abstract group used for comparison was the same as that for the transportation and food problems). They do not say whether they rotated schools and major fields on different problems; they say only that their rules "were expected to be consistent with the experiences of the subject population."

Unlike Van Duyne (1974), Yachanin & Tweney found no significant difference in responses between the thematic school group and the abstract group (school: 12%, abstract: 11%). This is true even if one considers only the affirmative (AA) problems. Yachanin & Tweney do not report having embedded their rules in a story context, social contract or otherwise.

School Problem Summary

One study found a content effect with the school problem and one study did not. Van Duyne (1974) found a content effect with the school problem when it was presented in a universal or standard conditional format. He found no content effect when the school problem was presented in a disjunctive or conjunctive
format, but there are a number of confounds that could have swamped an effect for these linguistic formats. Because his school problems were embedded in a context that made his rule part of a social contract rather than a simple descriptive relation, and because his subjects were, in essence, asked to "look for cheaters", it is difficult to tell whether the facilitation he found is due to the use of thematic content in general, or due to the social contract content that it has. Yachanin & Tweney (1982), who did not embed their problem in a social contract context, found no enhancement in logical performance with the school problem.

2.4 Social Contract Problems

A social contract specifies what two or more individuals intend to exchange. In a social contract, whether an individual receives a benefit is contingent upon his paying a cost or meeting a requirement. Chapter 5 provides a detailed account of the structure of social contracts; my purpose here is to give the reader an intuitive grasp of this structure, so I can review the relevant literature.

A social contract rule relates perceived benefits to perceived costs, expressing an exchange in which an individual is expected to pay a cost to an individual or group in order to be eligible to receive a benefit from that individual or group. "Cheating" is the violation of a social contract rule; more specifically, cheating is the failure to pay a cost to which you have obligated yourself by accepting a benefit from that individual or group. The other person would not have agreed to provide the benefit.
Cheating does not always correspond to logical falsification.

Consider the "Drinking Age Problem" (DAP; Griggs & Cox, 1982), pictured in Figure 2.1.

Figure 2.1

<table>
<thead>
<tr>
<th>Drinking Age Problem (DAP; adapted from Griggs &amp; Cox, 1982)</th>
</tr>
</thead>
<tbody>
<tr>
<td>In its crackdown against drunk drivers, Massachusetts law enforcement officials are revoking liquor licenses left and right. You are a bouncer in a Boston bar, and you'll lose your job unless you enforce the following law:</td>
</tr>
<tr>
<td>&quot;If a person is drinking beer, then he must be over 20 years old.&quot;</td>
</tr>
<tr>
<td>(If P then Q)</td>
</tr>
<tr>
<td>The cards below have information about four people sitting at a table in your bar. Each card represents one person. One side of a card tells what a person is drinking and the other side of the card tells that person's age.</td>
</tr>
<tr>
<td>Indicate only those card(s) you definitely need to turn over to see if any of these people are breaking this law.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>drinking beer</th>
<th>drinking coke</th>
<th>25 years old</th>
<th>16 years old</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P)</td>
<td>(not-P)</td>
<td>(Q)</td>
<td>(not-Q)</td>
</tr>
</tbody>
</table>

For American subjects who perceive drinking beer as a rationed benefit that can only be had by waiting until they have met an age requirement (the cost), the DAP expresses a social contract of the form:

"If you take the benefit, then you pay the cost."

This same rule would not express a social contract to subjects who do not think of the right to drink alcohol as an age-rationed privilege. I am told that in the USSR, people of any age can buy and drink alcohol: it is a "free good" with respect to age. For Russian subjects the DAP rule would be merely descriptive, relating a predisposition for drinking beer to advancing age.*

The transportation and food problems, as well as Yachanin & Tweney's school problem, were descriptive rules. Van Duyne's school problem was a prescriptive rule, but not a social contract

* Much as "If a person has a heart attack, then he must be over 20 years old" describes a relationship between advancing age and a predisposition to suffer heart attacks.
rule, because the benefit was not mentioned in the rule itself.

Figure 2.2 shows the structure of a Wason selection task that uses a social contract (SC) rule. Irrespective of logical category, a subject looking for potential cheaters should choose the "cost NOT paid" card (has he illicitly absconded with the benefit?) and the "benefit accepted" card (has he paid the required cost?).

Figure 2.2

<table>
<thead>
<tr>
<th>Structure of Social Contract (SC) Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is your job to enforce the following law:</td>
</tr>
<tr>
<td>Rule 1 — Standard Social Contract (STD-SC): “If you take the benefit, then you pay the cost.”</td>
</tr>
<tr>
<td>(If P then Q)</td>
</tr>
<tr>
<td>Rule 2 — Switched Social Contract (SWC-SC): “If you pay the cost, then you take the benefit.”</td>
</tr>
<tr>
<td>(If P then Q)</td>
</tr>
</tbody>
</table>

The cards below have information about four people. Each card represents one person. One side of a card tells whether a person accepted the benefit and the other side of the card tells whether that person paid the cost.

Indicate only those card(s) you definitely need to turn over to see if any of these people are breaking this law.

<table>
<thead>
<tr>
<th>Benefit Accepted</th>
<th>Benefit NOT Accepted</th>
<th>Cost Paid</th>
<th>Cost NOT Paid</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P)</td>
<td>(not-P)</td>
<td>(Q)</td>
<td>(not-Q)</td>
</tr>
</tbody>
</table>

Rule 1 — STD-SC: (P) (not-P) (Q) (not-Q)
Rule 2 — SWC-SC: (Q) (not-Q) (P) (not-P)

Whether looking for potential cheaters on a social contract produces logically falsifying card choices on the Wason selection task depends on where the costs and benefits to the potential cheater are located in the "If-then" structure of the rule.

A "standard" social contract (STD-SC) is one where the benefit to the potential cheater is located in the antecedent clause and the cost/requirement is located in the consequent clause. Rule 1 of Figure 2.2 and the DAP are STD-SCs. For a STD-SC, the "benefit accepted" card corresponds to the logical
category "P", and the "cost NOT paid" card corresponds to the logical category "not-Q".

A "switched" social contract (SWC-SC) is one where the locations of cost and benefit are switched -- the cost is in the antecedent clause and the benefit is in the consequent clause. Rule 2 of Figure 2.2 is a SWC-SC. For a SWC-SC, the "benefit accepted" card corresponds to the logical category "Q" and the "cost NOT paid" card corresponds to the logical category "not-P".

Consequently, looking for cheaters on a STD-SC produces the logically falsifying, 'P & not-Q' response, whereas looking for cheaters on a SWC-SC produces a logically incorrect, 'not-P & Q' response.

In the search for content effects on the Wason selection task, 16 experiments have tested rules whose content expresses a standard social contract -- the format for which 'P & not-Q' is the choice of a subject who is looking for potential cheaters. A substantial content effect has been found in every one of these experiments.

2.4.1 The Post Office Problem

A post office problem is a conditional rule expressing a postal regulation, for example, "If a letter weighs two ounces, it must have 44 cents postage." Whether a particular post office problem is a social contract or not depends on the subject population or the problem's context. It is a social contract problem:

1. if its constituent propositions are recognizable as costs and rationed benefits to the subject population, or
2. if the story context surrounding the problem defines the
constituent propositions as costs and rationed benefits. However, if its constituent propositions are arbitrary with respect to the subject population's understanding of costs and rationed benefits, then the same rule is merely descriptive.

For Americans, the post office rule, "If a letter is sealed, then it must have 20 cents postage", is either descriptive or prescriptive,* but not a social contract, because sealing an envelope is not considered a rationed benefit in our culture -- sealing is just something one always does, a free good. However, this same problem is a social contract problem for older British subjects, because in Britain before 1968 one could pay lower rates if one left the letter unsealed. In other words, the privacy gained by sealing a letter was a benefit that had to be paid for in that culture.

Johnson-Laird, Legrenzi, & Legrenzi, 1972

Johnson-Laird, Legrenzi, and Legrenzi (1972) formulated the first and most famous post office/SC problem. Their thematic rule was:

"If a letter is sealed, then it has a 50 lire stamp on it" Their subjects were British. In Britain, within the memory of their subjects, sealing an envelope was a benefit rationed by ability to pay; unsealed envelopes could be sent at a lower rate (this is no longer the case). Furthermore, the subjects were instructed to imagine they were postal workers looking for letters that "violate the rule". Thus, the problem asked them to

* Depending on whether the person saying the rule is simply making an observation about sealed letters (descriptive) or telling you a seemingly arbitrary postal regulation (prescriptive).
look for cheaters on a social contract. They were then shown the following display of five real envelopes with real lire stamps:

Figure 2.3 Thematic problem "card" display, Johnson-Laird et al., 1972

\[
\begin{array}{ccc}
\vdots & \vdots & \vdots \\
& & \\
\vdots & \vdots & \vdots \\
P & \text{not-}P & Q \\
\end{array}
\]

The envelopes are labeled with Ps and Qs for the reader's convenience; they were not so labeled in the actual experiment.

The logically correct answer is to choose 1) the sealed envelope (P), to see if the sender put too little postage, and 2) the 40 lire and no stamp envelopes (not-Q), which have too little postage to be eligible for sealing, to see if the sender illicitly sealed them. These are the same card choices a subject "looking for cheaters" would make.

Johnson-Laird et al. tested this rule in two different linguistic formats: "If P then Q" and "P only if Q" (e.g., "A letter is sealed only if it has a 50 lire stamp on it"). To reduce redundancy, one format used British stamps and the other use Italian stamps. Each of 24 subjects were given a total of four problems, one thematic and one abstract in each of the two linguistic formats. The order of problem presentation was randomized across subjects. The "cards" for the abstract problems were also real envelopes, with letters and numbers on the front and back. The rule reflected this by referring to the envelopes as "letters": "If a letter has an A on one side, then it has a 3 on the other side."
Ninety-two percent of the subjects gave the logically falsifying answer for at least one of the thematic problems, but only 29% gave this answer for at least one of the abstract problems. Seventy-one percent of the subjects got both thematic problems correct, but none of them got both abstract problems correct. There was no transfer from thematic problems to abstract problems, and when asked, only two of the 24 subjects realized that the logical structure of the thematic and abstract problems was similar. There were no effects due to linguistic format. Eighty-one percent of the 48 thematic problems (there were 2 per subject) were correctly solved, versus 15% of the abstract problems.

Golding, 1981

Golding (1981) conducted the Johnson-Laird, Legrenzi, & Legrenzi (1972) experiment using a population of British subjects ranging in age from 20 to 70 years. For subjects under 45 the post office problem was a prescriptive, non-SC problem, because the "lower rates for unsealed letters" rule was not in effect within their memory -- for them the constituent propositions were not recognizable as a cost and rationed benefit. Thus, there is no reason why these younger subjects would consider sealing an envelope a benefit that had to be paid for. However, this same
rule is a social contract problem for subjects over 45 who remembered the old regulation, because they would have recognized it as having a cost-benefit structure.

There was no difference in logical performance between the older and younger subjects on the abstract problem. However, 59% of the older subjects gave the logically correct, "look for cheaters" response to the post office problem, compared to only 9% of the younger subjects.

Golding herself framed the difference between the two groups in terms of familiarity rather than social contracts -- older subjects were more familiar with the rule than younger ones. In her view, older subjects perform better because they have direct experience with falsifying instances, and these instances are available in memory. In other words, older subjects do not reason about the rule, they simply recognize a pairing as falsifying if it occurs to them. There is nothing in her experiment to let one choose between her interpretation and a social contract interpretation.

Golding (1982)

Griggs & Cox (1982, Experiment 2) conducted Johnson-Laird et al.'s post office problem using 24 American college students. For American students, the post office rule is not a social contract, because its constituent propositions are not recognizable as a cost and rationed benefit (sealing an envelope is a free good in the U.S.). Griggs & Cox's procedure differed from Johnson-Laird et al. in only two ways: 1) they used American and Mexican instead of British and Italian stamps, and 2) for the abstract condition they made clear the fact that letters could
appear only on the front of an envelope, whereas numbers could appear only on the back. Because Johnson-Laird et al. did not say this, the blank envelope in the thematic condition was clearly an instance of not-Q, whereas the blank envelope in the abstract condition could be categorized as either not-P or not-Q. Griggs & Cox argue that this is a serious confound because "two not-Q instances in the thematic arrays highlight a selection that is critical for the correct solution" (p.412).*

There was no significant difference in performance between the thematic and abstract problems (thematic: 1/24 correct, abstract: 4/24 correct). Furthermore, the confound discussed above cannot have accounted for Johnson-Laird et al.'s results, because performance on both thematic and abstract problems was low in this experiment. If Griggs & Cox's highlighting explanation had been true, their adding the not-Q highlight to the abstract problems would have raised abstract performance to the high level elicited by Johnson-Laird et al.'s thematic problems, not lowered thematic performance to the level of abstract.

Van Duyne, 1976

Van Duyne (1976) briefly reports the results of an as yet unpublished study ("Semantics and reasoning", in preparation) comparing the performance of four independent groups of subjects

* Johnson-Laird et al. could counter that the ambiguity is irrelevant because their rule referred to "one side of the envelope" versus "the other side"; hence, to be safe, one should assume the blank was a potential not-Q. However, this is an extra, rather subtle inference that one need not make in the thematic condition. Thus, Griggs & Cox's objection would still stand, as the thematic problem alone contains two clear, unambiguous not-Q instances, and the issue would have to be decided empirically, as it was.
on the following four rules (one problem per subject):
1. an abstract rule
2. a descriptive post office rule using real envelopes as "cards": "If there is L.B. Mill on one side of the envelope then there is PRINTED PAPER REDUCED RATE on the other side". This rule's constituent propositions are not recognizable costs and rationed benefits to British subjects.
3. an SC post office rule using cards to represent envelopes: "If there is PRINTED PAPER REDUCED RATE on one side of the envelope then it must be left open". This is a real postal rule in Britain, where the experiment was run. To be eligible for the lower rates for printed matter, the contents of an envelope cannot contain any personal correspondence. Thus, the lack of privacy needed for the post office to make spot checks is the price you pay to be eligible for the benefit of saving money on mailings of printed material.
4. the same SC post office rule as in (3), using envelopes as "cards".

He reports the following percentage correct for the four problems: (1) abstract -- 19.27%, (2) descriptive post office -- 48.96%, (3) SC post office (cards) -- 86.98%, (4) SC post office (envelopes) 97.92%. He says these four form a "highly significant trend", but does not say what statistical tests he performed, or even the number of subjects in the study.

So many details are omitted from his description of this study that its relevance is difficult to ascertain. However, it seems unlikely that his sample sizes could be so small that the average performance on the social contract problems of 92.45% and 86.98% would not be statistically different from 19.27% for the
abstract problem, or even from 48.96% for the descriptive post office problem. Assuming that Van Duyne used the usual scoring system ('P and not-Q' is a "hit", any other answer is a "miss"), the minimum number of subjects required to get the number "97.92%" for group (4) is 48 (47 out of 48 correct). It is doubtful, however, that his group sizes were equal. If they were, the minimum number of subjects per group required for all four percentages to be possible numbers is 192. As this would mean his study used a total of 768 subjects, I presume he had unequal group sizes. But if group sizes did vary around 48, the differences between all four groups would be strongly significant. If this were true, his descriptive post office condition would also have shown enhanced performance over the abstract condition. But there are too many ifs to judge.

2.4.2 The Drinking Age Problem

The "Drinking Age Problem", or DAP, was developed by Griggs & Cox (1982). It is a rule relating eligibility to drink alcoholic beverages to age, for example, "If a person is drinking beer then he must be over 20 years old" (see Figure 2.1). In our culture, drinking alcohol is a benefit that one is eligible for only after having met an age requirement, so the drinking age problem qualifies as an SC problem for American subjects.

Griggs & Cox, 1982

Each of 40 subjects, undergraduates at the University of Florida, were tested on a DAP and an abstract problem. Their DAP was: "If a person is drinking beer, then the person must be over 19 years of age." Nineteen was the legal drinking age in Florida
at the time of the experiment. Before the DAP, each subject was told: "Imagine 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..." and so on. For both thematic and abstract problems subjects were instructed to select the card(s) necessary to "determine whether or not the people are violating the rule." Half the subjects were given the DAP first, half were given the abstract problem first.

Seventy-two percent of the subjects gave the logically falsifying answer, 'P & not-Q', for the DAP, but no subject gave this answer for the abstract problem. There was no transfer from a correctly solved DAP to the abstract problem. Also, the percentage correct for each problem was the same, whether it came first or second.

Cox & Griggs, 1982

Cox & Griggs, 1982 (Experiment 1) replicated their finding of a content effect with the DAP (thematic: 60%, abstract: 4%). Furthermore, although they also replicated the lack of transfer from a correct DAP to an abstract problem, they did find transfer from the DAP to the "apparel-color problem" (ACP): "If a person is wearing blue then the person must be over 19." The ACP is a "semi"-social contract problem: its consequent is a culturally typical cost for American subjects -- we are familiar with age-rationed privileges -- but the antecedent is not a rationed benefit. Interestingly enough, the ACP elicited a sizable content effect only when it followed the DAP, a full-fledged social contract.
Griggs & Cox, 1983

To see if thematic content could "protect" subjects against matching bias, Yachanin & Tweney had presented subjects with thematic and abstract rules whose components had been systematically negated (AA, AN, NA, NN; see section 2.1). None of their thematic rules, including the AA rules, had facilitated logical performance. Griggs & Cox decided to reopen this issue, because by 1983 two thematic problems had been found that could reliably elicit falsifying responses from American subjects: the DAP and the Sears problem (see section 2.4.3 below). Each subject was given only one problem to solve. There were four thematic (DAP) groups and four abstract groups: one group for each rule form (AA, AN, NA, NN), 20 subjects per group.

Although Griggs & Cox (Experiment 2) systematically negated the components of the DAP, they preserved their structures as standard social contracts. The four thematic rules were:

- **AA**: "If a person is drinking beer, then the person must be over 19"
- **AN**: "If a person is drinking beer, then the person must not be under 19"
- **NA**: "If a person is not abstaining, then the person must be over 19"
- **NN**: "If a person is not abstaining, then the person must not be under 19"

Even though the components have been systematically negated, in each case the "If" clause refers to the benefit in question (drinking beer) and the "then" clause refers to the requirement that must be met to be entitled to that benefit (being over 19).

Control groups solved one of four similarly negated forms of Wason's (1966) original abstract problem, "If a card has a vowel on one side, then it has an odd number on the other side." With this rule, subjects can recode negated phrases as affirmative phrases (e.g., "not odd" to "even"), just as they can for the
DAP. This is not possible with ordinary letter-number rules.\* 

Negated conditionals are frequently difficult to understand (Wason & Johnson-Laird, 1972, see 2.1 on Yachanin & Tweney, 1982). Yet, in spite of the fact that three of the thematic SC rules had negated components, every SC rule elicited more 'P \& not-Q' responses than its corresponding abstract rule (see Table 2.2).

<table>
<thead>
<tr>
<th></th>
<th>DAP</th>
<th>Abstract</th>
<th>Z-value</th>
<th>p-value</th>
<th>(\phi)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA:</td>
<td>70</td>
<td>10</td>
<td>3.87</td>
<td>&lt;.0001</td>
<td>.61</td>
</tr>
<tr>
<td>AN:</td>
<td>95</td>
<td>35</td>
<td>3.98</td>
<td>&lt;.00005</td>
<td>.63</td>
</tr>
<tr>
<td>NA:</td>
<td>75</td>
<td>10</td>
<td>4.16</td>
<td>&lt;.000025</td>
<td>.66</td>
</tr>
<tr>
<td>NN:</td>
<td>70</td>
<td>10</td>
<td>3.87</td>
<td>&lt;.00001</td>
<td>.61</td>
</tr>
</tbody>
</table>

Ave: 77.5 16.25 Z=7.76, p<.0000001 \(\phi\)=.61

Percentage 'P \& not-Q' responses, n=20 per cell. p values are one-tailed.

Note that falsification levels are highest for AN rules. This is always the case (Evans & Lynch, 1973; Manktelow & Evans, 1979; Reich & Ruth, 1982; Yachanin & Tweney, 1982), because the "matching response" for AN rules is also the falsifying response. For a rule like "If A then not-3", the subject who chooses 'A' and '3' has both matched the values mentioned in the rule and, by coincidence, falsified.

To sum up: four SC rules were tested, and all four elicited a robust content effect, even though three of these were phrased in unconventional ways (vis-a-vis the actual Florida Drinking Age Law), and had negated components.

\* for example, The "not-3" in "If A then not-3" cannot be recoded as a simple affirmative phrase.
2.4.3 The Sears Problem

The Sears problem was developed by D'Andrade (1981). It is set in a Sears store, and specifies the conditions under which a purchase must be authorized by the department manager. In real life, the purpose of such procedures is to make sure that customers or sales clerks do not cheat the store by "paying" for goods with a bad check, defunct or stolen credit cards, etc. Thus, it is a particularly nice example of a social contract problem.

D'Andrade (reported in Rumelhart & Norman, 1981)

D'Andrade's abstract and thematic problems both used prescriptive rules in the workplace. As part of your job, you (the subject) are supposed to make sure the rule has been followed. Half his subjects were given the abstract problem, half the thematic problem (this report does not say how many subjects were involved).

Both conditionals were embedded in the text of a story. In the abstract problem, "As part of your job as a label checker at Pica's Custom Label Factory, you have the job of making sure that all labels with a vowel printed on one side have an odd number printed on the other side." In the thematic problem, "As part of your job as an assistant at Sears, you have the job of checking sales receipts to make sure that any sale of over $30.00 has been approved by the section manager. (This is a rule of the store.)"

Both problems are set in the workplace, both use prescriptive rules and both invoke a "detective set" (Van Duyne, 1974) -- the subject is a "checker", looking for violations of a rule. But the Sears rule is a social contract whose violation
indicates theft -- a customer (or sales clerk) is not supposed to take something of great value without having paid its cost (as vouchsafed by the section manager). The abstract problem expresses a prescriptive rule, but not one whose terms are recognizable costs and benefits. Thirteen percent of subjects answered 'P & not-Q' for the abstract problem, whereas nearly 70% gave this answer for the social contract problem.

Griggs & Cox, 1983

Griggs & Cox's Experiment 2 included a condition identical to their DAP condition (described in 2.4.2), except that they tested systematically negated Sears problems. The abstract control groups for the Sears problems were the same as those for the DAP.

As with the DAP, Griggs & Cox systematically negated components of an AA Sears problem in a way that preserved their structures as standard social contracts. To do this, some of the negated components had to become quite long and complicated, using both explicit ("not") and implicit ("without") negatives to create what amounts to a double negative. The rules were:

AA: "If a purchase exceeds $30, then the receipt must have the signature of the department manager on the back."

AN: "If a purchase exceeds $30, then the receipt must not be without the signature of the department manager on the back."

NA: "If a purchase is not less than $30, then the receipt must have the signature of the department manager on the back."

NN: "If the [sic] purchase is not less than $30, then the receipt must not be without the signature of the department manager on the back."

For all four rule forms, the antecedent clause specifies that the rule pertains to high value purchases (large benefit), whereas the consequent clause specifies the authorization requirement
(vouchsafing that the high cost has been paid). Hence all four thematic problems are STD-SCs. The results are shown in Table 2.4.

Table 2.4 Griggs & Cox, 1983, Experiment 2 (Sears)

<table>
<thead>
<tr>
<th>Sears</th>
<th>Abstract</th>
<th>Z-score</th>
<th>p-value</th>
<th>phi</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA: 75</td>
<td>10</td>
<td>z=4.16, p&lt;.000025</td>
<td>phi=.66</td>
<td></td>
</tr>
<tr>
<td>AN: 70</td>
<td>35</td>
<td>z=2.22, p&lt;.013</td>
<td>phi=.35</td>
<td></td>
</tr>
<tr>
<td>NA: 60</td>
<td>10</td>
<td>z=3.31, p&lt;.0005</td>
<td>phi=.52</td>
<td></td>
</tr>
<tr>
<td>NN: 50</td>
<td>10</td>
<td>z=2.76, p&lt;.003</td>
<td>phi=.44</td>
<td></td>
</tr>
</tbody>
</table>

Ave: 63.75 16.25  z=6.13, p<.000001 phi=.48

Percentage of 'P & not-Q' responses, n=20 per group.
p values are one-tailed.

Despite abstruse, double negative circumlocutions, all four SC rules elicited a sizable and significant content effect. I would guess that the smaller percentages for the negated rules is due to their increasingly complex structure of negation.

2.4.4 Deformed Social Contract Rules

Deformed social contract rules have components that are recognizable as costs and rationed benefits. However, these components are arranged such that they violate the principles of social exchange, wherein one is obliged to pay a cost in order to be entitled to a benefit. Using deformed SC rules, one can see whether any prescriptive rule involving costs and benefits can elicit a robust and reliable content effect, or whether the costs and benefits must be arranged in the format of a standard social contract.

Deformed social contracts can be generated by systematically negating components of a STD-SC. For example:
AA: "If you take the benefit, then you must pay the cost."
AN: "If you take the benefit, then you must not pay the cost."
NA: "If you do not take the benefit, then you must pay the cost."
NN: "If you do not take the benefit, then you must not pay the cost."

The AN, NA, and NN rules violate the principles of social exchange.*

The ideal way to investigate deformed social contracts is to generate them from a rule that has no counterpart in the subject's experience. This can be done by embedding an unfamiliar rule in a story that defines its terms as costs and benefits, and arranging those terms in a format that violates the principles of social exchange. Deforming an SC rule that actually exists (like the DAP) introduces unfortunate demand characteristics. Subjects may assume the experimenter is testing their knowledge of the actual rule or law.** Alternatively, using a deformed version of an actual rule may "cue" the subject to reframe it as a proper (but different) social contract in a sci-fi world, and reason accordingly. The further a cost/benefit rule is from an explicit, real-life law, the better.

Unfortunately, the only experiments on deformed social contracts use the Sears problem and the DAP (Griggs & Cox, 1983, Experiment 1). The deformed DAPs use components of the Florida Drinking-Age Law, an explicit SC rule that actually exists. They are, therefore, especially vulnerable to the demand

* At first glance, it might seem that, although strange, the NN rule does not exactly violate rules of social exchange. However, in social exchange, a person who has not received a benefit is never prevented from paying the cost if he still wants to (see Chapter 5). This is, however, a more minor violation than the gross assymetries posed by the AN and NA rules.

** For example, a subject who assumed an experiment with deformed DAPs is actually testing her knowledge of the Florida Drinking Age Law would choose the "drinking beer" and "16 years old" cards, regardless of which logical categories they belong to.
characteristics noted above. Deformed Sears problems are somewhat less susceptible to the same interpretational difficulties because the AA Sears problem invokes more general SC principles, rather than an explicit, existing, law.

Griggs & Cox, 1983 (Experiment 1), systematically negated the components of the (AA) Sears problem and the DAP by inserting "not" or "does not" into the antecedent and consequent terms, regardless of how this affected the rule's status as a social contract. Hence, a phrase like "then you must pay the cost" would become "then you must not pay the cost." They compared these rules to ordinary letter-number abstract rules that had been similarly negated.

Deformations of the Sears Problem

The AA Sears problem does not exist as an actual rule or law. Rather, it invokes a common method for detecting potential cheaters in situations where money is being exchanged for goods. The rules that Griggs & Cox used are listed below. A translation into cost/benefit language is listed below each rule.

AA: "If a purchase exceeds $30, then the receipt must have the signature of the department manager on the back."

"If a customer takes a high value benefit, then we must make sure he has paid the cost.

AN: "If a purchase exceeds $30, then the receipt must not have the signature of the department manager on the back."

"If a customer takes a high value benefit, then we must not make sure he has paid the cost.

NA: "If a purchase does not exceed $30, then the receipt must have the signature of the department manager on the back."

"If a customer takes a low value benefit, then we must make sure he has paid the cost."
NN: "If a purchase does not exceed $30, then the receipt must not have the signature of the department manager on the back."

"If a customer takes a low value benefit, then we must not make sure he has paid the cost."

Griggs & Cox did not think of these problems in terms of social exchange. The reason they tested these deformations was to see whether the real (AA) rule would shine through, guiding the subject to a falsifying response. On this view, the AN, NA, and NN rules should all elicit content effects, and these should be of approximately the same size.

An alternative view is that only rules that instantiate standard social contracts elicit content effects. This leads to a different set of predictions, based on the principle that the deformed rule that most violates the principles of social exchange should elicit the weakest content effect.*

Hard cases make bad law — prediction would be easier had Griggs & Cox fabricated unfamiliar SC rules. However, social contract reasoning should elicit the following pattern of results for the AN, NA, and NN deformations of the Sears problem:

1. The NA rule violates the principles of social exchange the least. In real life, all purchases are benefits rationed according to ability to pay, even low cost purchases. The use of authorization signatures is a general method that can be applied no matter what the goods are, or how much they are worth: whether it is invoked for purchases which are over $30 or under $30 depends on how worried a manager is about the theft in each price category. It is certainly peculiar to require authorization of less valuable purchases,** but it does not do too much violence to the rule's status as STD-SC.

* Usually, no content effect at all. It depends on how easy it is to reframe a particular rule as a STD-SC.

** Especially if the NA rule leads one to infer that more valuable purchases do not require authorization. This inference is pragmatically reasonable, but logically invalid.
2. The AN rule violates the principles of social exchange the most. When a high value purchase has been made, it makes no sense to insist that a clerk refrain from making sure the cost has been paid. Furthermore, there is no easy way of reinterpreting the AN rule as a STD-SC. Thus the fact that the AN DAP suggests an SC, but is grossly deformed, should sow confusion, eliciting no content effect, or even a negative one, as compared to the AN abstract rule which so easily leads to falsification through matching.

3. The NN rule falls somewhere in between. If it said one need not check for authorization of low value purchases, rather than that one must not, it would not directly violate the principles of social exchange; in fact, it could easily be interpreted as an indirect way of stating the AA rule. However, the strange inclusion of must not might make subjects assume that both the NN and AA hold (as indicated by choosing all four cards). Pragmatic inference aside, in cost/benefit terms the NN rule has the same structure as the AN rule -- the violation is mitigated only by the fact that the antecedent of the NN rule represents a lower value benefit than that of the AN rule. Hence, it should not elicit a content effect.

To sum up: the social contract view predicts that the AN rule will elicit no content effect (or even a negative one), the NN rule will elicit no effect, and the NA rule will elicit a content effect that modest compared to that for the AA rule. The results are summarized in Table 2.5.

<table>
<thead>
<tr>
<th>Sears</th>
<th>Abstract</th>
<th>Z</th>
<th>p</th>
<th>phi</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA:</td>
<td>85</td>
<td>5</td>
<td>+5.09</td>
<td>&lt;.00000025</td>
</tr>
<tr>
<td>AN:</td>
<td>40</td>
<td>75</td>
<td>-2.24</td>
<td>&lt;.013</td>
</tr>
<tr>
<td>NA:</td>
<td>60</td>
<td>25</td>
<td>+2.24</td>
<td>&lt;.013</td>
</tr>
<tr>
<td>NN:</td>
<td>15</td>
<td>15</td>
<td>0</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Percentage 'P & not-Q' responses, \(n=20\) per group. \(p\) values are one-tailed.

As usual, the AA Sears problem elicited a large, significant content effect. The rest of the problems follow the social contract predictions. The NA Sears problem, which is a peculiar, but still proper, STD-SC, elicited a more modest content effect.
than the AA rule. The AN Sears problem, which grossly violates the principles of social exchange, actually elicited a negative content effect -- that is, subjects did better on the AN abstract problem than they did on the AN Sears problem. The NN Sears problem did not produce a content effect. Furthermore, more subjects chose all four cards on the NN Sears problem than on the NN abstract problem (40% v. 10%: Z=2.19, p<.015, phi=.35). This is what one would expect if subjects made the pragmatic inference that the NN Sears rule also implies the STD-SC AA Sears rule, but did not make the equivalent inference on the NN abstract rule.

This pattern of results is not predicted by Griggs & Cox's view -- in fact, they found the results puzzling. On their view, the AN and NN rules should have elicited content effects, just as the NA rule did. Griggs & Cox thought the negated rules would bring the AA Sears rule to mind and guide their inferences through "reasoning by analogy" to the logically falsifying choice. If subjects engaged in any reasoning by analogy to the AA rule, then the analogy must have been to the AA rule's structure as a standard social contract rather than to its structure as a logical conditional, because results for negated rules are best predicted by their cost-benefit structure.

Deformations of the DAP

The DAP does not invoke a general procedure for detecting potential cheaters, as does the Sears problem. Rather, it is a straightforward version of the Florida Drinking-Age Law, a specific, explicit law that was quite familiar to Griggs & Cox's University of Florida subjects. Thus, one can expect subjects' interpretations of deformed DAPs to be less flexible and more
vulnerable to demand characteristics than their interpretations of deformed Sears problems. The DAP rules used, and their translations into cost/benefit language, are listed below.

AA: "If a person is drinking beer, then the person must be over 19." "If you take the benefit, then you must pay the cost."

AN: "If a person is drinking beer, then the person must not be over 19" "If you take the benefit, then you must not pay the cost"

NA: "If a person is not drinking beer, then the person must be over 19" "If you do not take the benefit, then you must pay the cost"

NN: "If a person is not drinking beer, then the person must not be over 1 "If you do not take the benefit, then you must not pay the cost"

Griggs & Cox's view makes the same prediction as for the deformed Sears problems: the AN, NA, and NN rules should all elicit moderate content effects. As before, the social contract view has a different set of predictions:

1. The AN DAP severely violates the principles of social exchange. It should elicit no content effect, or even a negative content effect, for the same reasons that the AN Sears problem should.

2. The NA DAP is not so close to the AA DAP as the NA Sears problem is to its AA counterpart. Despite the negation, the antecedent of the NA Sears problem refers to a benefit that is rationed by ability to pay (purchases under $30), just as the antecedent of the AA Sears problem does. However, the antecedent of the AA DAP refers to a benefit that is rationed by age ("drinking beer"), whereas the antecedent of the NA DAP refers to a benefit that is not rationed by age ("drinking coke"). The NA Sears problem was merely a weak, or peculiar, STD-SC, whereas the NA DAP actually violates the principles of social exchange. Hence the NA DAP should not elicit a content effect.

However, there are reasons to believe the NA DAP will sow less confusion than the AN DAP. "Drinking coke" -- the NA DAP's negated component -- is not generally considered a liability by persons of any age. Thus, although "drinking coke" is not an age-rationed benefit, it is, at least, a benefit. In contrast, "must not be over 19" -- the AN DAP's negated component -- is not a cost/requirement for anything in our society. I can think of no benefit in our society that is open to adolescents but not to adults. Hence, the NA DAP violates the principles of social exchange somewhat less than the AN DAP. It is therefore less likely to elicit a negative
content effect than the AN DAP.*

3. Like the NN Sears problem, the NN DAP suggests a host of pragmatic inferences, which makes prediction difficult. The AA DAP is so well known that subjects undoubtedly realize it does not imply the NN DAP. Thus, it is unlikely that anyone would think that both the NN DAP and AA DAP hold, and choose all four cards. The NN DAP is certainly counter to experience, in that "drinking coke" is not a rationed benefit in our society, and "must not be over 19" is not a cost/requirement for anything. It may be so clearly counter to experience that it invites reframing as a proper STD-SC in a sci-fi world -- a world where adolescents take their revenge, and adults are not permitted the benefit of drinking coke, the drink of the adolescent. After all, the subjects were college students who were, or recently had been, below the legal drinking age. If some subjects made this sci-fi conversion, the NN DAP could elicit a modest content effect. This is why it is better to use unfamiliar rules -- existing laws invite too many interpretations to make sound prediction possible. Hard cases make bad law.

To sum up: The social contract view predicts no content effect (or a negative one) for the AN DAP, no effect for the NA DAP, and a question mark for the NN DAP.

As Table 2.6 shows, the results do not conform to Griggs & Cox's prediction of a content effect for all four problems -- the AN and NA DAPs did not elicit content effects at all. The pattern of results is best predicted by the social contract view.

The AA DAP, a STD-SC, elicited its usual hearty content effect. The AN and NA DAPs, which violate the principles of social exchange, did not elicit a content effect. Moreover, the AN DAP -- like the AN Sears problem -- was the only thematic

* The results of Evans & Lynch, 1973, also suggest that the NA rule might sow less confusion. Choice of the 'P' card is least influenced by matching. 'P' represents a true antecedent, and it is almost universally chosen due to a rudimentary understanding of logic or contingency. Thus, when a subject needs to reframe a bizarre rule to make sense of it, one might expect the subject to show greatest flexibility in re-interpreting the antecedent, and least flexibility in re-interpreting the consequent. If so, re-interpreting the AN DAP's negated consequent might be more difficult than re-interpreting the NA DAP's negated antecedent.
Table 2.6 Griggs & Cox, 1983, Exp 1 (DAP)*

<table>
<thead>
<tr>
<th>DAP</th>
<th>Abstract</th>
<th>DA Law</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA:</td>
<td>70</td>
<td>5</td>
</tr>
<tr>
<td>AN:</td>
<td>55</td>
<td>75</td>
</tr>
<tr>
<td>NA:</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>NN:</td>
<td>40</td>
<td>15</td>
</tr>
</tbody>
</table>

Columns 1 and 2: Percentage 'P & not-Q' responses, n=20 per group. The "DA Law" column shows the percentage of subjects who chose cards that indicate their knowledge of the actual Florida Drinking Age Law, irrespective of logical category. p values are one-tailed.

problem to elicit fewer falsifying responses than its abstract counterpart. Although the 20% difference between the AN DAP and its abstract counterpart falls 6 points short of a significant negative content effect, only the social contract view predicts that the AN data will move in this direction.

The NN DAP elicited a small content effect (phi=.28). This result neither confirms nor denies the social contract view, which makes no strong predictions about the NN DAP. However, there is evidence that knowledge of the actual drinking-age law prevented subjects from assuming that both the AA and NN DAPs hold: only 5% of subjects chose all four cards on the NN DAP, compared to 40% of subjects on the NN Sears problem. This difference is significant (z=2.65, p<.005, phi=.42).

Knowledge of the Florida law appears to have introduced other demand characteristics. The DA Law column of Table 2.6 shows the percentage of subjects who chose "drinking beer" and "16 years old", regardless of which logical categories these cards belonged to. These are the cards one would choose if looking for cheaters on the real drinking-age law. Because the negated DAP rules have quite obviously been yanked from an
existing SC law, subjects given a deformed DAP are far more likely to think the experimenter is testing their knowledge of the actual law than subjects given a deformed Sears problem.* The evidence supports this view. Significantly more subjects chose cards that would falsify the corresponding AA, STD-SC rule for negated DAPs than for negated Sears problems (23% v. 8%: Z=2.25, p<.013, phi=.21).

The results of Griggs & Cox's two experiments with deformed social contracts indicate that rules that violate the principles of social exchange do not elicit content effects. The closer a rule comes to the format of a standard social contract, the more likely it is to elicit a thematic content effect.

Social Contract Summary

Sixteen out of sixteen experiments comparing social contract rules to abstract rules have produced a robust content effect (Johnson-Laird et al., 1972; Van Duyne, 1976; D'Andrade, 1981; Golding, 1981; Griggs & Cox, 1982; Cox & Griggs, 1982; Griggs & Cox, 1983 (10 replications)). When a prescriptive post office problem was administered to cultural groups for whom the constituents were not recognizable as costs and rationed benefits -- that is, when subjects did not perceive the rule as a social contract -- no content effect was found (Golding, 1980; Griggs & Cox, 1982). Deformed social contracts -- rules that share constituents with proper social contracts but grossly violate the principles of social exchange -- do not elicit content effects (Griggs & Cox, 1983).

* only 1-2 subjects per negated Sears problem gave answers consistent with this interpretatation of the experiment.
General Summary of Chapter 2

Robust and replicable content effects are found only for rules that relate perceived benefits to perceived costs in the format of a standard social contract. No thematic rule that is not a social contract has ever produced a content effect that is both robust and replicable. For thematic content areas that do not express social contracts, either no content effect is found (the food problem), or there are at least as many studies that do not find content effects as there are studies that do (transportation and school problems). Moreover, most of the content effects reported for non-SC rules are either weak (Gilhooly & Falconer, 1974; Pollard, 1981), clouded by procedural difficulties (Bracewell & Hidi, 1974; Van Duyne, 1974), or have some earmarks of a social contract problem (Van Duyne, 1974). All told, for non-SC thematic problems, three experiments have produced a substantial content effect (transportation: Wason & Shapiro, 1971; Bracewell & Hidi, 1974; school: Van Duyne, 1974), two have produced a weak content effect (transportation: Gilhooly & Falconer, 1974; Pollard, 1981), and 14 have produced no content effect at all (transportation: Bracewell & Hidi, 1974; Manktelow & Evans, 1979; Yachanin & Tweney, 1982; Griggs & Cox, 1982. food: Manktelow & Evans, 1979 (4 experiments); Brown et al., 1982; Reich & Ruth, 1982; Yachanin & Tweney, 1982; school: Yachanin & Tweney, 1982. non-SC post office: Golding, 1980; Griggs & Cox, 1982). The few effects that were found did not replicate. In contrast, sixteen out of sixteen experiments with standard social contracts elicited substantial content effects. These include the Drinking Age Problem, the Post Office Problem, and the Sears
Problem. Deformed social contracts, which share constituents with standard social contracts but grossly violate the principles of social exchange, do not elicit content effects.

In this extensive literature, standard social contract rules are the only thematic content rules to elicit strong, replicable content effects on the Wason selection task.