INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600
THE ROLE OF EXPECTATION AS A MEDIATOR OF REASONING BEHAVIOUR

by

Jason Riis

A thesis submitted in conformity with the requirements for the degree of Master of Arts
Department of Human Development and Applied Psychology
Ontario Institute for Studies in Education of the University of Toronto

© Copyright by Jason Riis 1998
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-40668-7
Abstract

The role of expectation as a mediator of reasoning behaviour

Master of Arts 1998

Jason Riis

Department of Human Development and Applied Psychology

Ontario Institute for Studies in Education of the

University of Toronto

The Wason selection task (Wason, 1966) is a reasoning task where performance has often appeared to be suboptimal according to the norms of logic and hypothesis testing (Popper, 1959). The Theory of Relevance (Sperber and Wilson, 1986) has recently been used to explain performance on the selection task (Sperber, Cara & Girotto, 1995). The present study involves two experiments where variations of the selection task are used to test the Relevance account against two other current theories of reasoning -- the modularity account (Cummins, 1996a) and the theory of mental models (Johnson-Laird & Byrne, 1991). The results of both experiments favour the Relevance account. The account is also compared favourably with two other theories of reasoning -- the dual processing account (Evans & Over, 1996) and the subjective expected utility account (Kirby, 1994). Compatibility is noted with aspects of the general dual processing framework (Evans & Over, 1996; Sloman, 1996).
Acknowledgements

I would first like to thank David Olson for admitting me to the program and showing continued interest in my progress. I have benefited greatly from personal discussions with him as well as discussions in class and in research group meetings. On numerous occasions he pointed me in very useful directions with respect to work that I should read and issues that I should think about. David read various draft sections of this thesis, as well as the entire final draft, and offered substantial commentary. I often disagreed with his take on things, but I doubt if there is a more pleasant person in the world with whom to disagree.

The personal and intellectual debt that I have to my supervisor, Keith Stanovich, is not easily expressed. In our very first conversation Keith convinced me of the importance of studies of individual differences in cognition. It was not a problem that had concerned me before. This was just one of the many ideas that he exposed me to in our two years of work together. His enthusiasm for this field inspired me and has convinced me that full-time investigation of issues in higher cognition is a very worthwhile pursuit. Research efforts can guide policy efforts, as his work in the psychology of reading has shown. His professionalism and diligence have been a model of what it takes to engage in a successful and relevant research career. He has been an excellent teacher, role-model, and friend, and I look forward to following in his footsteps at the University of Michigan. He has helped me see what I should aim for, and has helped me develop the resources to achieve it.

My parents, Arlene and Fred Riis, also deserve considerable thanks. They have often provided me with much needed financial support, and have always provided me with encouragement to pursue my education. I'm very grateful for that.
Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>iv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vi</td>
</tr>
<tr>
<td>List of Appendices</td>
<td>vii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>The original task and confirmation bias</td>
<td>2</td>
</tr>
<tr>
<td>Content effects and deontic rules</td>
<td>5</td>
</tr>
<tr>
<td>Cummins' task variation for young children</td>
<td>12</td>
</tr>
<tr>
<td>Expectations and the pragmatics of conversation</td>
<td>14</td>
</tr>
<tr>
<td>EXPERIMENT 1</td>
<td>16</td>
</tr>
<tr>
<td>Method</td>
<td>17</td>
</tr>
<tr>
<td>Participants</td>
<td>17</td>
</tr>
<tr>
<td>Procedure</td>
<td>17</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>21</td>
</tr>
<tr>
<td>EXPERIMENT 2</td>
<td>24</td>
</tr>
<tr>
<td>Method</td>
<td>27</td>
</tr>
<tr>
<td>Participants</td>
<td>27</td>
</tr>
<tr>
<td>Procedure</td>
<td>27</td>
</tr>
</tbody>
</table>
Table of Contents (continued)

Results and Discussion ........................................................................................................ 28

GENERAL DISCUSSION ....................................................................................................... 32

Evans and Over’s (1996) dual processing account ......................................................... 34

Kirby’s (1994) Subjective Expected Utility account ..................................................... 44

Compatibility with dual processing accounts ................................................................. 52

Conclusion ......................................................................................................................... 55

Notes .................................................................................................................................. 59

References ......................................................................................................................... 60

Tables ................................................................................................................................ 65

Appendix ............................................................................................................................ 73
List of Tables

Table 1  Percentage of correct "Yard" selections in Experiment 1 .......... 65

Table 2  Number of subjects in each condition correctly performing the
several stages of Experiment 2 .......................................................... 66

Table 3  Number of subjects making particular selections in the
Falsification only and the Falsification and Confirmation conditions
for Experiment 2 ............................................................................. 67

Table 4  Number of subjects in each condition envisaging P & -Q, -P &
Q, both, or "other" as the combination that could falsify the
rule in Experiment 2 ......................................................................... 68

Table 5  Number of subjects in the Falsification and Confirmation
condition envisaging P & Q, -P & -Q, both, or "other" as the
combination that could occur if the rule were true in
Experiment 2 .................................................................................. 69

Table 6  Number of subjects identifying given combinations as potential
falsifying instances in the Falsification only condition ............... 70

Table 7  Number of subjects identifying given combinations as potential
falsifying and confirming instances ............................................. 71

Table 8  Number of subjects identifying given combinations as potential
falsifying and confirming instances ............................................. 72
# List of Appendix

<table>
<thead>
<tr>
<th>Appendix A: Other Results from Experiment 2</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>73</td>
</tr>
</tbody>
</table>
The role of expectation as a mediator of reasoning behaviour

Introduction

In the last several decades, a substantial amount of work in experimental psychology has been devoted to the investigation of human reasoning abilities. Several apparent limitations in these abilities have been described. Sub-optimal performance has been observed on tasks requiring reasoning in accordance with the normative rules of statistics (Kahneman & Tversky, 1979, 1982), economics (Larrick, Nisbett & Morgan, 1993), logic (Evans, Newstead & Byrne, 1993), and hypothesis testing (Wason, 1966; Kuhn, 1991, 1992). With respect to logic and hypothesis testing, a large proportion of studies have involved Wason’s four-card selection task (Wason 1966). Several general accounts of reasoning have developed largely in response to the results of selection task studies. Some accounts have argued that task performance points to specific biases or limitations in human reasoning (Evans & Lynch, 1973; Johnson-Laird & Byrne, 1991; Wason, 1966; Wason and Johnson-Laird, 1972). Others have suggested that successful performance on certain versions of the task indicates that human reasoning is domain specific, that is, the content of a problem, independent of its logic, will guide the way that people reason (Cheng & Holyoak, 1985; Cosmides, 1989; Cummins, 1996). A recent approach by Sperber, Cara, and Girotto (1995) suggests that task performance is governed by relevance considerations that apply to all kinds of communication, including communication in the form of instructions in a reasoning task. These authors have argued that when people engage in a reasoning task like the Wason selection task, they make spontaneous inferences to help them understand the problem. These inferences
generally lead the person to expect certain outcomes. In some situations these expectations will be optimal, and in others they won’t be.

In the present study, two experiments are performed testing their theory against competing accounts. In the first experiment the theory is compared to Cummins’ (1996a) modularity account, and in the second it is compared to Johnson-Laird & Byrne’s (1991) theory of mental models. In both cases, the Relevance account of Sperber and colleagues is supported. It is suggested that two other theories, the dual process account of Evans and Over (1996) and the subjective expected utility account of Kirby (1994), may also explain the present results, however, the Relevance account is still favoured for both empirical and theoretical reasons. In the final section it is suggested that the Relevance account fits quite well within a general dual processing framework (Evans and Over, 1996, Sloman, 1996).

The original task and confirmation bias

In the original form of the Wason selection task, subjects are told about a set of 4 cards each of which has a letter on one side and a number on the other side. Subjects are told that their task is to find out whether or not a conditional rule pertaining to the cards is true or false. The rule is “If a card has a vowel on one side, then it has an odd number on the other side”. Of the four cards, two are displaying their letter side and two are displaying the number side, so the subject would see, for example, cards showing A,D,4, and 7. Wason wanted to see the extent to which people would follow Popper’s (1959) falsification approach to hypothesis testing, which involved searching for counterexamples. Only an absence of counterexamples could demonstrate that the rule was true. Finding confirming instances does not bear on the truth of the rule because
even if several such instances are found, a single counterexample could still make the rule false. Accordingly, following Popper, subjects should select the vowel card (A) and the card with the odd number (7). Only a card with a combination of a vowel and an odd number could show the rule to be false, and these are the only cards that could display such a combination. Interestingly, however, only 5-10% of subjects tend to pick only these two cards (Evans, Newstead & Byrne, 1993). The most frequently selected combination is A and 4. Many subjects also pick the A card only. For the remainder of the paper I will adopt the notation widely used in the literature, which borrows from logic the convention of referring to a conditional statement as “If p then q”. With this “p/q” notation then, the p in Wason’s original example, just discussed, would be a card with a vowel on its letter side. A not-p card would have a consonant. Likewise, a q card would have an even number on its number side, while not-q denotes a card with an odd number. Using this notation then, the counterexample is a card with a p and an not-q. The modal selection task response of A and 4 is denoted p and q.

The original explanation of this result (Wason, 1966) was that subjects exhibited a confirmation bias, that is, rather than attempting to falsify the rule, they tried to confirm it, by finding instances of p&q combinations. Confirmation bias had already been suggested as the underlying thought process determining performance on another task developed by Wason (1960) – the 2-4-6 task. In this task subjects are told that the experimenter has a rule in mind that describes sequences of three numbers, an example of which is the sequence 2-4-6. After being given that example, subjects are then told that they may ask the experimenter two kinds of questions in order to determine which rule he has in mind. They may suggest a 3-number sequence and ask the experimenter if the
sequence is an instance of the rule, or they may suggest a rule and ask the experimenter if it is the one he has in mind. They may ask questions until they suggest the correct rule. The rule that the 2-4-6 triple is meant to exemplify is simply “any ascending sequence”, but most subjects take a long time before they suggest this simple rule. What seemed to happen is that subjects would form a hypothesis of the rule and then test it by asking the experimenter whether or not certain consistent sequences were examples of the rule. So, for example, subjects would often begin by asking the experimenter about such sequences as 6-8-10 or 14-16-18. The experimenter, of course, reports that these are indeed examples of the rule, as they are all ascending sequences. The subject would then typically ask if the rule governing the triples was “any sequence of consecutive even numbers”. The experimenter would reply that it was not and the subject would then, typically, generate another hypothesis (e.g. “the difference between consecutive numbers must be the same”) and ask about confirming or consistent instances (e.g. 3-7-11, 15-20-25). Such sequences are instances of the real rule, however the subject’s hypothesis is false. What subjects typically do not do is offer triples that would contradict their hypothesis. For the “equal difference” hypothesis, just mentioned, subjects do not test the hypothesis by asking the experimenter about a potentially falsifying triple where the difference between numbers is not equal, such as 1-3-8. Doing so would probably bring them to the correct solution sooner. The philosopher Karl Popper (1959) had been arguing that falsification was the optimal approach to scientific investigation. What Wason had claimed to show with his two tasks was that people do not tend to investigate the world in the manner prescribed by Popper, but rather, they seek to confirm their hypotheses instead of trying to falsify them.
Content effects and deontic rules

Some investigators commented on the artificiality of these two tasks and suggested that people should not be expected to reason effectively about such trivial abstract matters (Wetherick, 1971). To address this concern several versions of the Wason selection task were developed which involved a richer content than the original “abstract” numbers and letters. The rules described real world objects or events, and were not as arbitrary as the original vowel-even number rule. An early version inspired by these critiques was the postal rule problem by Johnson-Laird, Legrenzi, and Legrenzi (1972). British subjects were told that the post office was enforcing the following rule pertaining to any envelopes that were mailed: “If an envelope is sealed then it has a 25d stamp on it”. The subjects were then shown four envelopes, corresponding to the same logical cases as in the original task: true antecedent (p), false antecedent (not-p), true consequent (q), and false consequent (not-q). The p envelope was sealed and the postage could not be seen. The not-p envelope was unsealed, and again the postage could not be seen. The q and not-q cards had postage values of 25d and 15d respectively, and for both envelopes, it could not be seen whether or not they were sealed. Subjects were told that they should identify the envelopes that should be inspected in order to find out if there were any violations of the rule. Violations would consist of sealed envelopes having a 15d stamp, so the cards that should be inspected were the sealed envelope, and the one with the 15d stamp. Again these envelopes correspond to the p and not-q cases. Interestingly, with this version, a majority of subjects correctly inspected only these envelopes. Very few subjects made the “confirmatory” selections of p only, or of p and q.
The facilitation observed with the postal rule could not be explained merely by the fact that the rule was concrete rather than abstract in nature. Other concrete versions produced as many incorrect responses as the original abstract version. Manktelow and Evans (1979), for example, asked subjects to test the rule, "if I eat fish then I drink gin". Most subjects turned over the cards showing fish and gin, or just the fish card. Very few selected the “scotch” card which represented the not-q case, and was thus a potential violator. Also, Cheng and Holyoak (1985) found that American subjects did not tend to make the correct p and not-q selection with the postal rule that had produced a majority of correct selections among English subjects. They noted that a rule similar to the one employed in the task had actually been in effect in Britain when those original subjects were tested, while no such rule was in effect in the United States. It thus appeared that familiarity with the rule, beyond its mere concrete nature, was essential to lead to reliable reasoning performance. It turned out, however, that familiarity was not the only factor involved. If American subjects were given a rationale for the unfamiliar postal rule (i.e., sealed letters are usually personal and the post office charges more for personal mail), then they would tend to correctly select only the two potentially violating envelopes.

Two other problems which have reliably led to p\&not-q as the modal response are the Sears problem (Griggs & Cox, 1983) and the Drinking Age problem (Griggs & Cox, 1982). The rule for the Sears problem is “If a sale is for more than $100, then the receipt must be signed by the manager”. Subjects were asked to find out whether or not the rule is being violated. They tended to correctly select the p card (a receipt valued at $100) and the not-q card (a receipt not signed by the manager). For the Drinking Age problem, subjects are told that they are police officers and that they go into a bar to enforce the
following rule: "If a person is drinking beer, then they must be over 21 years of age".

Four people in the bar are represented by cards, which indicate what the person is drinking, and their age. Two of the cards reveal the person's beverage, "beer" and "coke", and hide their ages. Two hide the beverages and reveal the ages: "19 years" and "22 years". The p and not-q cases correspond to the "beer" and "19 years" cards, and indeed, it is only these two cards that a majority of subjects select.

What all three of these rules – the Postal rule, the Sears rule, and the Drinking Age rule – have in common is that they are deontic in nature. A deontic statement is one that involves a permission or an obligation (Manktelow & Over, 1991). Statements like "You must brush your teeth before bed" and "Library books must be returned on time" are examples of deontic statements. They are contrasted with indicative statements, which are just statements describing a particular state of affairs. Examples of indicative statements are, "All dogs are mammals" and "New York is a big city". The rule in Wason's original selection task was a conditional indicative statement of the from "if p then q". The Postal rule, the Sears rule, and the Drinking Age rule were, however, conditional deontic statements of the form "if an action is taken, then a precondition must be met".

The apparent facilitation of performance on selection tasks involving deontic rules led Cheng and Holyoak (1985) to suggest that a special cognitive process was governing performance on such tasks. They proposed that specific cognitive structures, which they called "pragmatic reasoning schemas", governed reasoning behaviour in different domains. These schemas were sets of relatively abstract, generalized production rules that were induced through a lifetime of experience with appropriate situations. The
schemas relevant to reasoning about deontic rules were the permission and obligation schemas. The permission schema, for example, allows people to reason about regulations, "...imposed typically by an authority to achieve some social purpose" (p.398). It is hypothesized to consist of the following four production rules:

Rule 1: If the action is to be taken, then the precondition must be satisfied.
Rule 2: If the action is not to be taken, then the precondition need not be satisfied.
Rule 3: If the precondition is satisfied, then the action may be taken.
Rule 4: If the precondition is not satisfied, then the action must not be taken.

Cheng and Holyoak argue that most of the rules that have produced facilitation on the Wason selection task have been permission rules in the form of Rule 1. Such a rule is thought to trigger the other rules. When examining the array of cards presented in the selection task, Rule 1 would lead to the selection of the p card (corresponding to "the action is taken") to see if not-q ("the precondition is not satisfied") is on the back. Rules 2 and 3 would not lead to the selection of any cards, because their consequences only describe possibilities, not certainties, and hence no single consequence could bear on their application. Finally, the not-q card would be chosen, because Rule 4 of the schema forbids the action under such circumstances. Application of these rules, triggered by the permission rule predicts the selection of the p and not-q cards. The Drinking Age problem, the Sears problem, and the Postal problem, are all permission rules where there
is either an explicit or implicit social purpose. On all of these problems, p and not–q was the modal response.

This schema account has been challenged by Cosmides (1989). Like Cheng and Holyoak (1985) she suggests that facilitated performance is the result of the operation of a specific cognitive mechanism, but the mechanism which she proposes is different in two important ways. First, Cosmides argues that the mechanism is innate. She calls it a Darwinian algorithm to suggest that it evolved as an adaptation – in this case, an adaptation to life in a society of individuals. The second difference pertains to the kind of situations under which the algorithm operates. The schema account suggests that people will reason effectively in any situation involving a deontic statement – either a permission or an obligation. Cosmides argues that effective reasoning will not be observed over so wide a range of situations. Only some deontic rules will induce effective reasoning, and these are rules which involve what she calls “social exchanges”. A social exchange is a situation governed by a deontic rule where a person takes a benefit for a cost. Not all deontic rules will involve cost/benefit elements. The cognitive algorithm that she proposes puts people on the alert for others who accept a benefit without suffering the cost. It will not be activated in deontic situations where costs and benefits are not involved. Cosmides points out that all of the selection task rules that Cheng and Holyoak have called permission rules, are, more specifically, cost-benefit rules. So while they do take the general form “if an action is to be taken, then a precondition must be met”, the action is always something that gives a person a benefit, and the precondition is always something that involves the person suffering a cost. For example, the postal rule (“If an envelope is sealed then it has a 25d stamp on it”),
describes how a person may get the benefit of mailing a more secure letter by paying the
cost of extra postage. For Cosmides, a permission rule that has a clear social purpose will
not by itself lead to optimal cheater detection behaviour as Cheng and Holyoak
suggested, but rather, the rule must be understood as one involving a social exchange of
benefit and cost.

To test this, she compared performance on a \textit{deontic} rule in two contexts. In one
context the rule was given a non-cost-benefit social purpose, and in the other it was
presented a social exchange involving a benefit and a cost. One rule that she used was "If
a student is to be assigned to Grover High School, then that student must live in Grover
City". A student violating the rule would attend the school but not live in Grover City
(i.e. p and not-q). Students attending another high school (not-p), Hanover High, and
students living in Grover City (q) would not be violating this rule. In the non-cost-benefit
context, the rule was presented as having the social purpose of ensuring that the Board of
Education can maintain the statistics that it needs to assign the right number of teachers
to each school. In this case a violation would be an administrative error by the secretary
assigning the students. In the cost-benefit context, however, a violation would involve a
student or parent seeking to gain an advantage without suffering the cost. In this
scenario, the subject is told that Grover High School is considerably better than Hanover
High, but that Grover City is a less desirable place to live than the town of Hanover.

Going to Grover High School is thus a benefit that may be obtained only by incurring the
cost of living in the less desirable Grover City. When presented in this latter context,
75\% of subjects correctly selected only the p and not-q cards, while only 30\% of subjects
did so in the former, non-cost-benefit context. In both contexts, a (deontic) permission
rule with a social purpose was tested. The fact that only the more specific social purpose involving cost and benefit facilitated performance argues against the existence of the more general permission schema and in favour of the more specific Darwinian social exchange algorithm.

Cummins (1996a) has recently developed a position that combines elements of Cosmides' (1989) Darwinian algorithm account and Cheng and Holyoak’s (1985) schema account. She agrees with Cosmides that the relevant cognitive mechanism is innate, but she sides with Cheng and Holyoak with respect to the range of situations over which it operates. She argues that the Darwinian algorithm that Cosmides describes is too narrow. She points out that while it is true that awareness of cost-benefit considerations would have been important to early hominids, other deontic considerations would have been important as well. She emphasizes the fact that humans and earlier hominids (and apes before them) lived in hierarchical societies. In such societies, resources are often allocated based on social rank. Lower order males, for example, may not simply mate with whichever females will have them. Higher males demand priority. Individuals with a higher rank also have priority over shared food, and they do not necessarily have to incur greater costs to receive this benefit. There is a subtle difference here. For Cummins, a person will incur a great cost by violating a deontic rule. But this is a different kind of cost than that described by Cosmides. For Cosmides, the cost is incurred not by violating the rule, but rather, by complying with it. The cost is the price of the benefit one receives, so there is a trade-off. For Cummins this trade-off needn’t be present. A tribal chief, for example, may be entitled to various benefits due to his rank, and it’s not obvious that having that rank involves any serious costs. To survive in such a
hierarchy, one must be keenly aware of what one is permitted and obliged to do. Specifically one must be very aware of what constitutes a rule violation. Cummins suggests that cognitive processes that could handle such considerations would be most advantageous. Accordingly, she argues that the deontic-indicative distinction is hard-wired into the human cognitive architecture, and that a cheater detection strategy is invoked when a person reasons about a deontic rule. A cheater is someone who violates a deontic rule, regardless of what costs or benefits may have been involved.

Empirical evidence supporting this broader distinction (as opposed to Cosmides' narrower "cost-benefit versus non-cost-benefit" distinction) comes from more work by Cheng and Holyoak (1989). They found that a majority of subjects correctly selected the p and not-q cards when asked which cards could violate the rule "If you go out at night, then you must tie a piece of volcanic rock to your ankle". This is a deontic rule that does not seem to lend itself to cost/benefit considerations.

**Cummins task variation for young children**

While Cummins largely agrees with Cheng and Holyoak about the range of the phenomenon, she opposes their suggestion that it is the result of induced schemas that develop through an individual’s experience. She has recently offered another example of a non-cost/benefit deontic rule that leads to correct performance, although this time her subjects were 3- and 4-year old children (Cummins, 1996b). She argues that the demonstration of the deontic/indicative distinction in young children is evidence that the distinction is innate.

In her study, Cummins showed the children some toy mice who lived in a cardboard box, which was their house. The children could see that some of the mice
were in the house and some were outside the house, playing in the yard. They were then told that while the mice all looked the same, some were in fact different, because some of them squeaked and some did not. The children were handed two mice, only one of which squeaked when squeezed. The child squeezed each mouse in turn. They were then told that the squeaky mice make a lot of noise when they go into the backyard, and that the noise attracts the neighbourhood cat. A toy cat was then introduced and was shown to chase the mice around the yard. Accordingly, the children are told that it is not safe outside for the squeaky mice, it is only safe for the quiet mice.

Both the deontic and indicative versions shared that introductory story. For the indicative condition, the children were then introduced to a toy character, Minnie Mouse. They were told that Minnie likes to tell kids things and that they should listen carefully because they will have to find out if what Minnie says is right or wrong. Minnie then tells the child that “All the squeaky mice are in the house”. The child is then asked which mice should be squeezed to find out if Minnie is wrong; the ones in the house or the ones in the yard. The correct response is to squeeze the mice in the yard. Minnie is wrong only if there is a squeaky mouse in the yard. It doesn’t matter which mice are in the house.

For the deontic version, the Minnie Mouse character is introduced as the Queen Mouse who makes important rules that everyone must follow. The children are told to listen carefully because the Queen is going to make a rule and they will have to make sure that no one disobeys. The Queen then says that, “All the squeaky mice must be in the house”. The child is then asked which mice must be squeezed to make sure that none of the mice are disobeying; the ones in the house or the ones in the yard. Again the
correct response is to squeeze the mice in the yard, because only squeaky mice in the yard could be disobeying the Queen. All mice are allowed to be in the house.

In a between subjects design, Cummins found that more 4-year-old children correctly squeezed the mice in the yard for the deontic version, where 80% did so, than for the indicative version, where 37% did so. For 3-year olds, the percentages were, similarly, 68% and 32%, respectively. She takes the result to be evidence that the deontic-indicative distinction is a primitive one in the human cognitive architecture.

Expectations and the pragmatics of conversation

Recent work by Sperber, Cara, and Girotto (1995) has called into question the significance of the deontic/indicative distinction in the adult literature. They have suggested that the respective rule types do not, inherently, invoke a particular reasoning strategy. Rather, the strategy that will be adopted is determined by the expectations that the subject develops in a given situation. Subjects will come to expect particular outcomes, and this expectation will depend partly on how likely these outcomes seem. One task that they used involved a scenario where a prince had just dismantled all welfare programs in his country, including medical insurance and unemployment insurance. He defends his extreme free market policy by suggesting that the programs are not needed because, “if a person in the country is of working age, then that person has a job.” The participant is asked to imagine herself as a journalist who must investigate the prince’s claim. In response, when presented with cards representing different people in the country, the majority of subjects correctly selected cards which could potential falsify the rule, that is, cards that could represent a person who is of working age but who is unemployed. Note that this is an indicative statement, not a deontic one. With most
indicative rules, subjects seem to search for confirming instances. Sperber and colleagues argue that they search for falsifying instances here because they doubt the truth of the rule and expect counterexamples. The notion of an unemployed person of working age would also be quite salient to the college-aged subjects, many of whom may themselves be concerned about finding work. Upon reading the rule, a representation of that particular counterexample would be readily generated.

This analysis can be applied to Cummins' task. In the deontic case, the Queen says that “All of the squeaky mice must stay in the house”. This implies that there is some possibility that the squeaky mice will be outside. In conversation, deontic rules are discussed generally in situations where violation of the rule is at issue. Children are often told that they may not do certain things at just the time when they were about to do them, or shortly after they have done them. When presented with the deontic rule in the Cummins task then, children may expect violations, and hence will be more inclined to correctly check the mice in the yard. This is not the case for indicative statements. Such statements do not carry an inherent suggestion that it would be reasonable to think that there were counterexamples. When Minnie tells the children that “All the squeaky mice are in the house” the children have little reason to doubt that she is telling the truth. Accordingly they will expect that the squeaky mice are in the house. In squeezing the mice in the house, they are testing their expectation that some of them will be squeaky. It is not being suggested here that this is the correct way to test the rule. Clearly it is not. The point is that the indicative scenario made salient the likelihood that there were squeaky mice in the house. On the other hand, the deontic scenario made salient the possibility that there were squeaky mice in the yard. This salience is what forms the
basis of the child’s expectation, and it is the expectation that leads to the choice of which mice to check.

Now, what I want to suggest is that the particular expectations that are formed, are not determined uniquely by the deontic or indicative nature of the rule, although that is a factor, as the two statements do imply different states of affairs. What they imply are different intentions on the part of the speaker. That someone forbids a certain behaviour implies that they are concerned that the behaviour might occur, and they wish to prevent it. A listener detects the concern and makes inferences accordingly. This concern does not usually exist when someone utters an indicative statement, although there are certainly situations when it can.

Sometimes when someone utters an indicative statement, the listener will infer that they are concerned with the possibility that it is not true. In such situations we may be much more inclined to doubt what the person says. The examples that Sperber and his colleagues used involved situations where a speaker’s indicative statement was a denial in response to an accusation. In such cases they found that their adult subjects were much more likely to search for falsifying evidence of the statement than they were if the statement was not made in response to an accusation.

Experiment 1

In Cummins’ task, I suggest, the deontic version invited considerably more doubt than did the indicative version. In the present study the aim is to equate that doubt. Cummins herself seems to have been sensitive to this point as she notes that in her indicative version, Minnie was introduced as an unreliable character, as the experimenter
tells the children that “Minnie likes to tell kids things”. This however does not go very far in suggesting that what she says should be doubted. In the present study, using Cummins’ task, we manipulate the indicative scenario in such a way that Minnie’s indicative statement that “All the squeaky mice are in the house” is taken as a denial. Minnie is introduced as a character named “Meannie Mouse”, who is very mean, and collaborates with the cat. The children are told that Meannie doesn’t like the squeaky mice and that she helps the cat to catch them. In this situation, we anticipate that the children would doubt that what Meannie says is true. If the children make note of Meannie’s motivations, then they may take her indicative statement to be a denial – a denial which should be doubted. Accordingly, they should expect counterexamples in the form of squeaky mice in the yard. We thus expect more of them to check the yard than will do so in Cummins’ original indicative version. The frequency of yard inspection should be similar to that observed using her deontic condition.

Method

Subjects

Subjects were 53 four year olds ranging in age from 4 years, 0 months to 4 years, 11 months. The children were drawn from several day care centres in the city of Toronto. SES varied from lower middle to upper middle class.

Procedure

Children were randomly assigned to one of three conditions, and were tested individually. Each condition consisted of a version of Cummins’ task. Two conditions were replications of her original deontic and original indicative versions. The third condition was a modification of her indicative version.
For all conditions, the task was preceded by a brief warm-up, to ensure that the children were comfortable talking to the experimenter. The child was shown a 2 dollar coin and asked to identify what appeared in the pictures on either side of the coin. On one side is a picture of a polar bear and on the other side is the picture of a queen (a young Queen Elizabeth). All children were able to identify the pictures (Approximations were accepted). The child was then asked to stand up, drop the coin to the floor from head height, and to identify which side of the coin was showing. This was repeated a second time.

The child was then seated in front of a table where the cardboard house and the mice were displayed. Four toy mice were inside the box, and four were beside the box in the area later described as the yard. In all three conditions, children were told the following introductory story:

<Name of child>, I'd like you to play a pretend game with me. I'd like you to pretend that this is a house, and these mice [pointing to mice] live there. That's their home. Some of the mice have gone out into the backyard to play, see? They're having a good time playing with the ball. So these mice are in the house [pointing] and these mice are in the yard playing [pointing]. At night, everybody goes in the house, and they close the door (door is affixed) so that they're nice and safe and nothing can get them. (door removed). Now, <name of child>, would you do me a quick favour? Would you point to the mice that are in the house? And would you point to the mice that are in the backyard?
I have something else very interesting to tell you about these mice. See how they all look exactly alike? But they’re really different because some of them squeak (demo) and some of them are quiet (demo). Squeeze this mouse and you’ll see that it make a lot of noise (child squeezes). Now squeeze this one (child squeezes). See? It’s really quiet. It doesn’t make any noise at all.

I have something else interesting to tell you. Sometimes in the evening, the mice like to go out in the backyard to play. But when the squeaky mice play, they get really excited and they start to squeak, like this (squeeze repeatedly). And then you know what happens? The neighbourhood cat hears that squeaking, and he comes running, and pounces, and chases the mice all around (cat appears and chases mouse). So it’s not safe for the squeaky mice outside. It’s only safe for the quiet mice.

The remainder of the protocol was then different for each condition. The three protocols went as follows:

**Cummins’ Original Deontic.**

Now the Queen Mouse heard about this (Minnie appears). The Queen Mouse is their mom, and she makes important rules that everyone has to follow. So let’s listen carefully because she’s going to make a rule and we’re going to have to make sure nobody disobeys. Queen Mouse: “Oh, I’m so worried about the squeaky mice because of that cat, so I’m going
to make a rule, and the rule is all squeaky mice must stay in the house. Yes, all squeaky mice have to stay in the house.”

She said all the squeaky mice must stay in the house, so let’s make sure nobody is breaking that rule, OK? Let’s make sure no one is disobeying the Queen. Which mice should we check, those that are in the house, or those that are in the backyard playing?

**Cummins’ Original Indicative.**

Now I’d like you to meet someone. This is Minnie mouse. (Minnie appears). Minnie likes to tell kids things so let’s listen carefully because we’re going to have to figure out whether what she says is right or wrong. Minnie Mouse: “Hi <name of child>! I know something about the mice. Uh huh! I know that it’s not safe outside for the squeaky mice because of that cat, so all the squeaky mice are in the house! Yes, all the squeaky mice are in the house!”

Now, I wonder if she could be wrong about that. Let’s find out. Which mice should we squeeze to find out if she’s wrong, the mice that are in the house, or the mice that are in the backyard playing?

**Modified Indicative.**

Now I’d like you to meet someone. This is Meannie Mouse (Minnie appears). She’s very mean and likes to trick people. She doesn’t like the squeaky mice because they make a lot of noise, so
she likes it when the cat chases them. She even helps the cat catch the squeaky mice.

I want to ask her a question: Hey Meannie, it’s not safe for the squeaky mice to be in the yard because of the cat, so are all the squeaky mice safely in the house? Now, Listen carefully to what she says because we're going to have to make sure she's not wrong.

Meannie: (in a suspicious tone of voice) “He-he...Oh yes...Don’t worry...all the squeaky mice are in the house, yes all the squeaky mice are in the house.”

Hmm. I wonder if Meannie might be lying to us about that.

She said all the squeaky mice are in the house, but I wonder if maybe that’s wrong. Let’s find out if it’s wrong. Which mice should we check? Those in the house, or those in the backyard playing?

For all conditions, the order of query (inside first or outside first) was counterbalanced between subjects, as was the relative position of the house and the yard (left and right, or right and left).

The child was required first to point and state verbally which mice had to be squeezed to test the rule, and was then allowed to actually squeeze the mice.

Results and Discussion

The percentage of children who indicated that the mice in the backyard had to be checked is shown in Table 1. In the original deontic condition 71% of the children checked the yard. Thirty-three percent did so in the original indicative condition, while
56% did so in the modified indicative condition. Chi square tests revealed that the greater number of correct yard selections in the original deontic condition than in the original indicative condition was significant, $X^2 (1) = 3.82, p < .01$. No other differences were significant. It should be noted, however, that the sample size was fairly modest. Were the trend to maintain and show significant differences among all three groups, then this would call into question Cummins’ (1996a) suggestion that indicative rules invoke, willy nilly, a strategy of confirmation, while deontic rules invoke a strategy of violation detection. Were the modified indicative condition to lead to more correct (violation detection) selection than the original indicative version, then this would suggest that indicative reasoning is a function of expectation, since the modified version was designed to invoke the expectation that violations would occur.

Recent work with adults has suggested that deontic rules maybe prone to similar manipulations. Kirby (1994) found that if people can be made to expect that a deontic rule will be complied with, then they will be less likely to employ a falsification strategy when testing for such compliance. Using the Drinking Age version of the selection task he found that most people would fail to select the not-q card if that card showed the age of the person being 12 years old rather than 19 years old. Recall that the rule to be tested was, “If a person is drinking beer then that person must be at least 21 years of age”. A 19-year old drinking beer would violate this rule, as would a 12-year old. Kirby suggests, however, that fewer people select the not-q card when it represents a 12-year old because they do not think that such a young child would be likely to violate the rule. In other words, they did not expect a violation.
Falsification in deontic reasoning may then, as with indicative reasoning, be a function of a person’s expectation. A similar test could be done with young children using a variation of Cummins’ deontic version. Two suggestions come to mind. One would be to indicate to the children that the mice were very well behaved and always obeyed the Queen. Also, instead of telling children to make sure that no one disobeys, the instruction could be to find out if everyone was obeying. Framing the instruction in such a positive manner may reduce the expectation of violation. If children expect compliance they will expect to find squeaky mice in the house. Even though the presence of squeaky mice in the house will not demonstrate full compliance, the salience of the expectation may be impossible to ignore.

The present account, following Sperber, Cara and Girotto (1995), views the thought process underlying selection behaviour on tasks like the one described here, not as any kind of conscious reasoning process, but rather, as one of spontaneous inference. Such inferences are not guided by reflective thought. They are unconscious, produced perhaps by complex associations, rather than by the application of mental rules. What is interesting is that the spontaneous inferences or expectations that adults form, seem similar to those formed by very young children.

The next step in the research program would be to understand more specifically just which aspects of particular situations inform this process of spontaneous inference. One possibility is that adults and children may give priority to the desires of other people as they consider a situation. In the present study, falsification behaviour in the modified indicative case may have resulted from the children focusing on Meannie Mouse’s desire to catch the squeaky mice. This desire requires squeaky mice in the yard, so the
possibility of their presence there may have been inferred accordingly. In the deontic case, the declaration of the rule may have led to a focus on the desires of those whom the rule governs, namely the squeaky mice. The children may have noted the desire of the mice to play in the yard, and thus may have inferred the possibility of their presence there. Finally in the original indicative case, the children may have focused on the desires of Minnie and the rest of the mice to stay safe. Without the statement of the deontic rule, the desire of the mice to play in the yard may have been less apparent, and the desire for safety, leading to the inference of squeaky mice in the house, may have prevailed.

Bartsch and Wellman (1995) have recently provided theoretical and empirical evidence for the priority given to desires in young children’s theory of mind. The relation of such work to the present task could be explored by asking children what the various characters desired and seeing if their reports of such desires predicted their selection behaviour. This could be a direction for future work.

Experiment 2

While many researchers have focused their studies and theories on the effects of manipulating content with the Wason selection task, others have continued to focus on the original abstract task. A domain general account of the selection task has been offered by Johnson-Laird and Byrne (1991) with the theory of mental models (MMT). MMT proposes that reasoners construct models of premises, and reach conclusions based on the content of these models. Models represent possible cases (in the world) which are allowed by the premises, and also cases which are not allowed. Initially, all of the cases
may not be represented. Some cases will need to be fleshed out. On this account, errors in reasoning result from a failure to flesh out all the cases correctly, leaving relevant cases unrepresented. The nature of the conclusions is thus a function of the discretely represented cases.

Recently a prediction of MMT has been tested by Green and Larking (1996). The theory predicts that if subjects correctly represent the case which cannot occur, then they will select only the cards that could reveal such instances, namely the p and \(-q\) cards. To test this, Green and Larking devised a procedure that would assess whether or not a subject had correctly represented the counterexample before the subject was asked to make the card selection. This procedure involved two stages. They called the first stage “envisaging”. Subjects were told of the nature of the cards, as in the original task (i.e. a letter on one side and a number on the other side) and were then told of the possible rule that “if there is a vowel on one side of the card then there is an even number on the other side”. They are then asked which combination of a letter and a number should not occur on a card if the rule is true. A correct response involved writing down only an instance of a vowel and an odd number (i.e. \(p\&\neg q\)). Subjects were then shown the standard 4 card array, with one card each showing a vowel, a consonant, an even number, and an odd number, representing the \(p\), \(\neg p\), \(q\), and \(\neg q\) cases respectively. The second part of the procedure involved “identifying” the card or cards on which the “envisaged” type of combination could occur. Obviously, this will just be the \(p\) card and the \(\neg q\) card, but while Green and Larking found that 96% of subjects were able to correctly envisage the combination that should not occur, only 60% (48/77) were able to correctly identify only the vowel and odd number cards as the ones that could contain such a combination. Of
interest here are the 60% who envisaged and identified correctly. Green and Larking maintain that all of these subjects would have a mental model of the counterexample, and hence, according to MMT they should select only those cards when asked to find out whether the rule is true or false. At very least, they should recognize the importance of the not-q card and select it. It turned out that the prediction was largely borne out, as 90% of subjects who correctly envisioned and identified, went on to select not-q when given a standard selection task instruction. While a small number of subjects did not select not-q, the fact that the vast majority did, seems to support the claim of MMT.

The result is, however, also consistent with the theory of Sperber, Cara and Girotto (1995) which predicts that subjects select cards that could reveal outcomes that they expect, regardless of how they bear on the truth of the rule. It could be argued that the externalization procedure focuses attention on the counterexample by asking questions about it. It thus encourages participants to expect the counterexample, so that those subjects who represent it, will select it.

To test this possibility, the present study modified Green and Larking’s task so as to reduce the expectation of the counterexample while still ensuring that a large number of subjects accurately represented it. This was done by adding a second externalization task, which involved envisaging and identifying a combination that could occur if the rule were true. Of interest, were participants who represented the p&q combination as the one that could occur, and represented the p&not-q combination as the one that could not occur. These participants should be less likely to expect a counter-example than would be subjects who perform the standard procedure alone (and, hence, only form the
representation of p&not-q), simply because the counterexample expectation would have to compete with the expectation of confirming p&q cases.

Method

Subjects

One hundred and eighty-eight subjects participated in the study and received credit for coursework. Subjects ranged in age from 18-23 years, and all were students at a medium sized university in West Virginia.

Procedure

After filling out a demographics questionnaire, subjects were given either Green and Larking's (1996) original version of the task ("falsification only") or the modified version ("falsification and confirmation").

Both conditions were presented as a paper and pencil exercise. For both conditions, participants were told about a set of cards, each of which had a letter on one side and a number on the other side. They were told that their task was to find out whether a particular rule was true or false of four cards that were on display. The rule was "If there is a vowel on one side of the card then there is an even number on the other side", and the four cards on display had A, D, 4, and 7 on their face-up sides. This is consistent with the standard presentation of Wason's selection task -- the cards represented the p, not-p, q, and not-q cases respectively. Following presentation of the rule, subjects were asked to follow a series of instructions in order to find out whether the rule was true or false.

For the falsification only condition there were three such instructions, which were exactly the same as those used by Green and Larking. First, subjects were asked "Which
combination of a letter and a number should not occur if the rule is true?”. Green and Larking called this the “envisaging” task. For the next instruction, pictures of the 4 cards were shown again (A,D, 4, 7) and participants were asked “on which card or cards could this type of combination occur?” Green and Larking called this the “identification” task. The final instruction was the standard selection task question, “which card or cards do you need to turn over in order to see if the rule is true or false?”. The description of the task and the above instructions were all presented on a single sheet.

For the falsification and confirmation condition there were five questions to be answered following the task description. Again the description and the questions all appeared on a single sheet. The first two questions were the same as for the falsification only condition. These questions involved, respectively, “envisaging” a falsifying combination, and “identifying” the cards on which this combination could occur. The third and fourth questions involved envisaging a combination that could occur if the rule were true, and then identifying the cards on which this combination could occur. Specifically, the third question asked “which combination of a letter and a number could occur on a card if the rule is true?”. The fourth question asked, “on which card or cards could this type of combination occur?”. The fifth and final question was the same selection question given in the falsification only condition – “which card or cards do you need to turn over in order to see if the rule is true or false?”.

Results and Discussion

Of interest was only a subset of subjects (Other results from this experiment are given in Appendix A). For the falsification only condition, only those subjects who correctly envisaged and identified the counterexample were of interest. Only these
subjects could be said to have mentally represented the counterexample according to Green and Larking's operational definition. So only the performance of these subjects would be relevant to the prediction of MMT. Correct envisaging involved indicating that a combination of a vowel and an odd number (p and not-q) could not occur if the conditional rule were true. Subjects' responses were coded as correct if they gave an example of such a combination (e.g., "A and 7") or if they described this type of combination (e.g., "a vowel with an odd number"). Of the 93 subjects in the falsification only condition, 69, or 74%, correctly envisaged the counterexample. This was lower than the 96% who did so in the Green and Larking study, however, it is compatible with the percentages from Green's (1995) earlier study where only 80% correctly envisaged the counterexample. Correct identification involved correctly envisaging the counterexample and then identifying on which of the four cards such a combination could occur. This involved circling both the A card (p) and the 7 card (not-q), and only those two cards. Of the 69 participants who correctly envisaged, 36, or 52%, correctly identified only the p and not-q cards. So a total of 36 out of 93 participants, or 38%, could be said to have mentally represented the counterexample.

Mental Models Theory would predict that all, or at least the great majority of this subset would go on to select the not-q card when asked which cards needed to be turned over in order to find out if the rule was true or false. In fact, only 58% did so. This is lower than the percentage that did so in the Green and Larking study, where 90% of the specified subset selected not-q, however, it is considerably higher than the percentage that typically select not-q in the standard selection task, so this result alone does not clearly refute MMT.
The main test of MMT involved comparing this 58% who selected not-q after correctly envisaging and identifying the counterexample with those who selected not-q after correctly envisaging and identifying both the counterexample and a consistent combination. This latter group was a subset of the subjects in the falsification and confirmation condition. Such subjects were first required to envisage and identify the counterexample. Seventy-six percent (n=73 out of 95) correctly envisaged and 46% (n=33) of those also correctly identified. These numbers are, as expected, similar to those from the falsification only condition. Here, then, 33 out of 95, or 34%, correctly envisaged and identified, and hence, mentally represented the counterexample. These subjects were also asked to envisage and identify a combination that could occur if the rule were true. Now, while only the p and not-q combination can falsify the rule, there are three combinations that can occur if the rule is true (p & q; not-p & q; not-p & not-q). Of interest, however, were only subjects who envisaged and identified the p & q combination or the not-p & q combination. Subjects who envisaged and identified the not-p & not-q combination would be correct, however they would have a second mental model involving the not-q card (in addition to the p and not-q model of the counterexample). If they selected not-q when testing the rule, it would thus not be clear if they did so because they thought it might reveal a counterexample or because they thought it would reveal an instance of not-p & not-q. We would thus be unable to test the prediction that representation of the counterexample leads to selection of the not-q card, because it may have been due to representation of the confirming not-p and not-q combination. If a subject selects not-q after envisaging and identifying p & q as the confirming case, however, then only the representation of the counterexample could be
responsible. So the relevant subset in this condition included only subjects who had envisaged and identified the counterexample and who envisaged and identified p & q as the combination that could occur if the rule were true. There were 21 such subjects, or 22% of the 95 subjects in this condition. Of interest was the percentage of these subjects who correctly selected the not-q card when asked which cards needed to be turned over in order to find out if the rule is true or false. Only 24%, or 5 out of 21, did so. Since all 21 subjects had a mental model of the counterexample, MMT would predict that a large majority of them would have gone on to select not-q. This clearly did not occur. The 24% who did so in this condition was significantly less that the 58% (of relevant subjects) who did so in the falsification only condition, $X^2(1) = 6.4$, $p<.01$. These results are summarized in Table 2.

Our main hypothesis was supported by the data. Subjects who were able to identify cards on which a counterexample could appear were less likely to select not-q when testing the conditional statement if prior to selection they were also asked to identify the cards on which a combination consistent with the rule could occur. Our suggestion was that subjects identifying both potential falsifying and potential confirming instances would be more likely to expect both kinds of instances, and that these expectations would compete. For many subjects the expectation of p&q cases would be stronger so they would search only for such instances. They would thus be less influenced by their representation of the counterexample. Subjects who did not have to identify the potentially confirming instances, but only the falsifying instances, would tend to expect only these instances and thus select cards on which they could appear. Contrary to MMT then, the correct selection of cards is thus not only a function of a subject's
success in representing the counterexample, because both groups of subjects represented
the counterexample. Something else is involved – expectation of the different
combinations.

Another possibility is that selection results from demand characteristics involved
in asking subjects to perform the identification stage. The subject may infer that since the
experimenter is bringing their attention to the counterexample, the counterexample must
be important and they should accordingly select cards that may reveal it. Similarly they
may assume that confirming instances are important if the experimenter draws attention
to them.

On the basis of the present experiment alone, it is not easy to rule out either
possibility. A combination of the two factors, expectation of a particular card
combination (likelihood), and importance of a particular card combination (value) may
interact. Such an interpretation would be consistent with a number of recent
interpretation of selection task results. A number of investigators have modeled card
selection on a decision theoretic basis, whereby appeal is made to the likelihood and
value of selections.

General Discussion

The present results argue against the modularity account offered by Cummins
(1996a) and the basic Mental Models Theory (Johnson-Laird & Byrne, 1991). In
Experiment 1, a variation of the selection task was presented to children in three
conditions. Each involved a different scenario. One condition involved a deontic rule,
and two involved an indicative rule. One of the indicative conditions involved a scenario
that was modified from Cummins’ (1996b) original indicative scenario in such a way as to lead children to expect violations of the indicative rule. There was a trend suggesting that this condition led to fewer correct selections (searching for violations) than the deontic version, but also to more correct selections than the original indicative version. The correct selection rate for this modified indicative version was 56%, calling into question Cummins’ modularity account which suggests that children will test indicative rules using a confirmation strategy. The trend here was for a majority to use, rather, what she would call a violation detection strategy.

In Experiment 2, it was found that with the abstract selection task, selections could not be based solely on the subject’s having formed mental models of particular cards and cases. It was demonstrated that forming a mental model of the falsifying combination does not necessarily lead to selection of the not-q card. This was seen when a group of subjects in the falsification and confirmation condition was less likely to select the not-q card than was a group in the falsification only condition, even though both groups had showed evidence of having successfully represented the p and not-q combination as an outcome that should not occur. It seems that in addition to forming a mental model of the card combination, a subject must also feel that such outcomes are either likely or important.

It has been argued that the results of both experiments can be explained by the Relevance account offered by Sperber, Cara, and Girotto (1995). These authors argued that the modal selections on the selection task are the product of spontaneous inferences. With these inferences, the subject forms an expectation for certain outcomes, and then goes about searching for such outcomes, assuming that their presence or absence will
reveal the truth or falsity of the rule. Two other theories, however, may also be able to handle the present results. These are the dual process account of Evans and Over (1996) and the Subjective Expected Utility account of Kirby (1994). I will discuss these two accounts in turn, and argue that on theoretical and empirical grounds, both are inferior to the account of Sperber and colleagues.

**Evans and Over’s (1996) dual process account**

Evans and Over (1996) have elaborated previous versions of Evans’ (1984, 1993) heuristic/analytic account. This account views thinking as consisting of dichotomous processes. The heuristic process is automatic and largely preconscious, while the analytic process is more explicit and intentional. The heuristic process guides the analytic process in that it creates the representation over which the analytic processes will operate — that is, it determines the focus of attention. Heuristic processes are thus a severe constraint on analytic processing. Heuristics are necessary, however, because without them, analytic processing would take far too long and would thus be of little help to an organism that had to make quick decisions (Oaksford & Chater, 1995).

The representation that the heuristic process will settle on is the one that is the most “relevant”. This use of relevant is not the same as the notion of relevance used by Sperber, Cara and Girotto (1995). Here the notion is more closely related to Tversky and Kahneman’s (1973) notion of “availability”. The notion of availability was originally used to explain the errors that people make in frequency judgement. For example, people tend to assume that more words start with the letter “r” than have “r” as the third letter. The availability account suggests that the human memory system organizes words according to the first letter so when we try to think of examples of words starting with “r”
we can think of very many compared to the number we can think of having “r” as the third letter. This differential is what is used to judge the general frequency of the two kinds of words. The judgement is, in fact, wrong as there are more words having “r” as the third letter than as the first, even though their availability in memory suggests the opposite. The point is that an unconscious process cues the memory of the different kinds of words and thus forms the representation on which the judgement is made. This much is common to the Over and Evans relevance account. But while availability has generally been used to explain perceived frequency (or event likelihood) the Evans and Over notion of relevance also requires that an outcome be related to the person’s personal goals. An outcome is said to be of high utility if is perceived to be important to these personal goals. Evans and Over suggest that heuristic processes assess likelihood and utility. Outcomes of sufficient likelihood and utility will become the focus of attention. Evans and Over put it as follows:

An available model will become relevant and thus the focus of attention provided that it has high goal relevance, which is effectively the same thing as a high expected cost or benefit. (p. 81)

They discuss some findings by Manktelow and Over (1991) to illustrate how personal goals determine the focus of attention. These authors presented subjects with a selection task using the rule “If you tidy your room then you may go out to play”. This is pragmatically equivalent to “If you go out to play, then you must have tidied your room” where “going out to play” is the p case and “tidied room” is the q case. The subjects were to imagine that the rule was spoken by a mother to her son. In a situation like this
the personal goals of the mother and son would be different so each would be concerned about different outcomes. The mother’s priority is that the room be cleaned while the son’s priority is that he gets to play outside. The mother will thus be concerned about a rule violation whereby the son goes outside but does not tidy his room. In selection task notation, this would be a p and not-q instance. The son on the other hand has the priority of going outside. A violation of the rule for him would be a situation where he had tidied his room but was not permitted to go outside. This would be a not-p and q instance.

Manktelow and Over instructed half of their subjects to assume the role of the mother when they did the task, and the other half to assume the role of the son. Each was asked which cards needed to be turned over to see if the rule was being violated. What they found was that the majority of subjects adopting the mother’s perspective selected the p and not-q cards while those adopting the son’s perspective selected the not-p and q cards. These two modal selections are the ones that would reveal the outcomes that each party would be most concerned with. Similar findings have been reported by Gigernezer and Hug (1992) and Politzer and Nguyen-Xuan (1992). Evans and Over (1996) take the results to suggest that subjects internalized the personal goals of the characters in the stories and focused on outcomes that were most pertinent to those personal goals.

For Evans and Over (1996), then, relevance just determines what will become the focus of attention. In the selection tasks presented by Manktelow and Over (1991) a particular outcome became the focus of attention. Once attended to, they argue, the subject will think of nothing else, and will merely rationalize the importance of that particular outcome. This does not leave much of a role for thought or inference. Sperber, Cara, and Girotto (1995), on the other hand, hold that subjects make spontaneous
inferences of consequences, and the role of relevance is to inform them as to whether or not the inference that they have made is plausible under the circumstances.

The notion of relevance discussed by Sperber, Cara and Girotto (1995) was developed by Sperber and Wilson (1986). Sperber and Wilson followed Grice’s (1975) account of communication whereby conversation (including written instructions of the kind used in reasoning tasks) is facilitated by an assumption of co-operativeness among the participants. This means that for every act of communication there is a presumption that the information contained in the utterance is relevant to the addressee. The addressee may not find relevance, or she may doubt that the speaker really intended any, but in general, she will at least try to find relevance in the utterance. If the given information does not seem directly relevant, then the addressee will make inferences. This happens often. The following example, taken from Spreber, Cara, and Girotto (1995, p.51) offers an illustration:

Peter: Do you want to go to the party at the Smiths’?

Mary: They came to our party.

Mary has not given Peter the exact information that he wanted, namely, a direct indication of whether or not she’d like to go to the party. However, as the addressee Peter assumes that her reply is relevant, and in the context is thus able to infer that she would like to (or, at least, feels obliged to) go to the party. So if the uttered information is not directly relevant, then inferences will be made. This happens in the lab setting as well. A subject may not be able to find much relevance in the experimenter’s rule, “If p then q”. She will thus make inferences until relevance is achieved. The question is, then, what inferences will she make. Here, Sperber, Cara, and Girotto (1995) offer two
principles that guide the inference process. These are the principles of *effect* and *effort*. The principle of effect suggests that an inference is more likely to be made if it reveals new and precise information. Recall that for the standard abstract selection task, Sperber and colleagues suggest that subjects spontaneously infer from the conditional “if p then q” rule that p and q instances should exist. This inference satisfies the principle of effect because it suggest a directly testable consequence of the original utterance (even though it is invalid). It thus provides the subject with a course of action, in terms of choosing cards (i.e., the p card and the q card), which is exactly what the subject was after. It is in this sense that the inference provides new and precise information.

The principle of *effort* is satisfied by this inference as well. This principle suggests that inferences which require minimal effort are more likely to be made than ones requiring more effort. This is satisfied by the inference that “p and q instances will exist” partly because this inference does not involve the representation of a negation. Negations, Sperber and colleagues point out generally require more effort to process (Horn, 1989). Inferring that “p and not-q instances should not exist” requires representation of two negations and is thus going to be more effortful, even though it is correct.

This is different form the account offered by Evans and Over (1996), in that relevance always regards inferences, which are in the form of propositions. For Evans and Over, relevance regards mere representations of objects – relevance makes certain representations the focus of attention without necessarily making any assertion about the objects. Evans and Over’s relevance can focus attention on an object but it won’t necessarily lead to a belief about the object. For Sperber and colleagues, there is always
a belief. Relevance does not merely focus attention on the p and q case, but it leads to the
inference (the belief) that such cases ought to exist.

Another difference between the two accounts is that Evans and Over view the
heuristic processes underlying “facilitated” versions of the selection task (e.g. the
Drinking Age problem, or the “tidied room” problem) as being quite different from those
involved in the standard abstract version. Recall that for the ‘tidied room” problem
(Manktelow and Over, 1991) the personal goals that the subjects adopted for the task, by
assuming a particular role, determined the utility of various outcomes which in turn
determined relevance. Likelihood of an outcome, they suggest, could also determine
relevance.

They argue, however, that with the standard abstract task, the subject simply does
not have any information about the likelihood of the possible outcomes (i.e. letter/number
combinations on the cards), nor any personal goals to which any of the outcomes are
relevant. Accordingly, they argue that for such tasks, instead of the pragmatic relevance
of likelihood/utility factors, linguistic relevance operates at the heuristic stage to direct
the focus of attention. The focus of attention is no longer the outcome having the highest
expected utility, but rather, the focus is merely on the properties that are mentioned in the
rule. So where pragmatic relevance puts a combination of features in focus (e.g.
instances of p and not-q), linguistic relevance puts only individual features in focus (i.e.
the faces of particular cards, rather than combinations). For the original task, the focus is
on the p and the q cards. Note that this is very different from the Sperber and colleagues
account which suggests that subjects expect p and q combinations. With Evans and
Over’s linguistic relevance, however, subjects do not have any expectations. They do not
select the p and q cards because they expect p and q combinations, but rather, they select them because heuristic processes automatically draw their attention, separately, to the p feature and to the q feature. This is what Evans and Lynch (1973) originally called the “matching bias” — subjects just select the cards that are mentioned in the rule.

So Evans and Over propose two kinds of heuristic processes, those that direct attention to combinations of features through pragmatic relevance, and those that direct attention only to single features, through linguistic relevance. The latter are invoked if pragmatic cues of likelihood and utility are not present. They offer two empirical findings to demonstrate the existence of the second, linguistic, heuristics. The first is the finding that if subjects are presented with a negated conditional of the form “if p then not q” then they tend to correctly select the p and the q cards (Evans 1972). This is the correct selection response because only combinations of p and q could falsify this negated conditional rule. Evans and Over, (following Evans, 1972) suggest that subjects are not making an insightful selection, but rather, are merely making the matching response. This is more easily illustrated by using an example. If the rule was “if a card has a vowel then it does not have an even number” then the correct selection would be the vowel and even cards, and these are the two cards that are mentioned in the rule.

Although the rule specifies that vowels do not have evens, “evens” are mentioned. Odd numbers are not.

The other support that Evans and Over claim for the linguistic relevance heuristic comes from a recent study by Evans (1996). Evans developed a new methodology whereby subjects were presented with a standard selection task on a computer screen, and asked to point with a mouse at any card that they were currently considering. They were
to *click* the mouse only when they were sure that they wanted to select the particular card.

What Evans found was that for a given card, subjects spent significantly more time inspecting that card if they ended up selecting it than if they did not select it. Subjects spent very little time inspecting cards that they did not end up selecting. Evans and Over (1996, Evans 1996) argue that this supports the suggestion that subjects make selections based on the way that cues of relevance focus attention on certain cards. For the abstract selection task, the theory predicts that attention would be focused on the p and q cards. The look-time data showed that subjects did tend to look at these cards longer than they looked at the not-p and not-q cards. Taking look-time as an indicator of a subject's focus of attention, they conclude that subjects tended to choose exactly those cards that they were focused on. This was the case not only for the abstract version of the task, but also with versions that led to predominantly correct selections. On these versions, subjects looked more at the p and not-q cards, and these were the cards that most ended up selecting.

Once heuristic processes draw attention to certain cards, Evans and Over suggest that analytic processes then take over. These processes then allow the subject to rationalize their selection of the cards that they are focused on. So the analytic processes are not involved in the decision process -- only in the post-decision rationalization. Sperber, Cara, and Girotto (1995) point out that it is strange that the two processes would have such a relationship. It would not be very adaptive for an organism to evolve with two reasoning processes that always come up with the same answer when it comes to directing behaviour. One would expect at least some divergence. As we will see later in the discussion, Sperber and colleagues do favour a dual processing account, but not one
where initial decisions are merely rationalized. Rather, they suggest, if initial decisions do not seem plausible, a more elaborate reasoning process (which they call "meta-inference") will be invoked to approach the problem. I will discuss this suggestion further in the final section.

I should first, however, point out that the relevance account of Sperber, Cara, and Girotto (1995) seems able to handle both the findings from, the look-time studies, and from work with negated conditionals. Under their account, subjects spontaneously infer consequences of the selection task rule, and then go about deciding upon which cards those combinations could occur. With the look-time task on a standard abstract rule, most subjects will infer that p and q instances should occur. They will, accordingly, focus on the p and q cards, and the look time could be spent checking that these cards could indeed reveal the expected instances of p and q. Under this account subjects would not be expected to spend much time looking at cards that could not reveal the expected outcomes, so the results of Evans' (1996) study are quite consistent.

As for the findings with the negated conditional, Sperber and colleagues argue that the presence of the negation in the actual conditional statement alerts the subject that the statement is denying the existence of something. The inference made, then, is more likely to be relevant if it is also a denial — a denial of the existence of certain combinations (i.e., instances of p and q). Note that the positive conditional "if p then q" implies that p and not-q cases shouldn’t exist, however, it is not itself a denial. It is taken as a positive statement so inferences from it will tend to be positive as well. The so-called 'matching bias" can thus be explained by appealing to the kind of inference that is more likely to be make in the face of differently valenced conditionals. This is an
alternative to acknowledging the operation of Evans and Over’s (1996) “linguistic relevance cues” which merely focus a subject’s attention and say nothing of the kind of inferences that a subject might be making.

I have argued that the results supporting Evans and Over’s (1996) dual processing theory are quite consistent with the Relevance account of Sperber, Cara, and Girotto (1995). I would further suggest that the latter account is to be favoured for reasons of parsimony. Evans and Over propose two distinct kinds of relevance cues at the heuristic stage – linguistic and pragmatic. Linguistic cues operate when there is not sufficient information available relating to the subject’s personal goals. It is not clear however, that even in informationally impoverished situations (such as the abstract selection task) that subjects’ personal goals would not be involved. A subject would surely still have the goal of trying to understand the task and would thus be trying to make sense of the information that is presented. This, as Sperber and colleagues point out, will involve making inferences. They suggest that the processes that lead to these inferences are the same regardless of the content of the task. The subject assumes that the task information is relevant, and then makes inferences that are consistent with that assumption. Constraints of effort and effect guide these inferences. For abstract tasks subjects tend to infer that p and q instances should occur. For some other versions, such as the Drinking Age problem, subjects will tend to infer that p and not-q instances should not occur. In both cases the subject has inferred a testable consequence. The process that governs the relevance considerations is certainly complex (and detailed in Sperber & Wilson, 1986), but it is a process that guides reasoning on all versions of the selection task. The account is thus more unified than the dual process account of Evans and Over.
Kirby's (1994) Subj ective Expected Utility account

Kirby's (1994) theory is similar to that of Evans and Over (1996) in that it relies on the decision theoretic notions of likelihood and utility. He suggests that subjects will make selections based on the subjective expected utility (SEU) of each card. Cards with a higher SEU will tend to get selected. The term “expected utility” is used because a choice will only be made if there is some likelihood that a valuable outcome will occur. Even if an outcome is extremely valuable (of high utility), the opportunity to achieve it may be forgone, if the likelihood that the outcome will obtain under the chosen opportunity is very small. In such a situation, the maximum possible utility is very high, but the expected utility is low, because the person does not expect the high utility outcome to obtain.

Kirby applies this analysis to the selection task. Over several experiments he manipulates both the utility and likelihood of falsifying (p&not-q) instances. One of these results was reported above: For the Drinking Age problem, subjects were less likely to select the not-q card if it revealed an underage 12-year old than if it revealed an underage 19-year old. Both cards have the potential to reveal the high utility outcome of an underage drinker, but the 12-year old is selected less often because a 12-year old is less likely to be trying to drink alcohol than is a 19-year old. The expected utility is thus less.

Kirby also argues that there is always a cost (i.e., a negative utility) involved in selecting cards that do not reveal cheaters. In the Drinking Age scenario, subjects are servers who would not want to bother customers by unnecessarily asking them for identification if they were drinking beer, or asking them what they were drinking if they
were underage. So selection of a card is not only determined by the likelihood of the falsifier, but also by the respective utilities of the possible outcomes, which, in the case of the not-q card, are a positive utility for finding a cheater, and a negative utility for a false alarm (i.e. checking which beverage the underage person is drinking and finding that it is a Coke). The negative utility for the false alarm would seem to be minimal in this case, but Kirby attempts to manipulate this. Subjects are either told to avoid offending customers by checking too much ("Don’t Check"), or to be sure not to miss anyone who could be violating the rule ("Don’t Miss"), at risk of getting fired. It turned out that more subjects in the latter ("Don’t Miss") condition selected the not-q card than did so than in the former ("Don’t Check") condition. This was the case when the not-q card displayed the person’s age being 19 years, and also when the age was 12 years and when it was 4 years. For subjects told not to miss any violations, the utility of finding a violation was very high. For subjects told to avoid unnecessarily bothering customers, the utility of avoiding a false alarm (i.e. avoiding finding not-p on the back of not-q) was high. Kirby argues that these different utilities were responsible for different rates of not-q selection.

Kirby’s account involves a two stage selection process. The first is an inference stage and the second is a choice stage. Both stages, he argues, involve content-independent processes, but can still be influenced by content. Content, for example, can influence the way that a person interprets a statement. In the selection task, certain problem content may lead to a biconditional interpretation of the rule (that is, the “if p then q” statement is taken to mean “p if and only if q”). But given that interpretation, the inferential process is content-independent and involves representation of combinations that can and cannot occur in light of the rule. But given these inferences, choice
processes must still determine which cards will be selected. The choice process is also content-independent, and cards which have the highest subjective expected utility will be chosen. These will be the cards that have the right combination of likelihood and utility. For these cards the SEU of selecting the card will be greater than the SEU of not selecting it (or, mathematically, SEU[choosing A] > SEU [not choosing A]). The likelihood and utility of a card will be influenced by the content of the problem, such as the “Don’t Check” or “Don’t Miss” instructions described above. But given the likelihoods and utilities, the choice process itself is content independent.

Unlike the account of Evans and Over (1996), Kirby says that SEU could account for performance on the abstract selection task as well. In this sense it shares the parsimony advantage of the Sperber, Cara, and Girotto (1995) account where spontaneous inferences based on relevance considerations explained the performance of most subjects, regardless of the content of the particular selection task.

Kirby argues that SEU applies to the abstract selection task as follows. For Wason’s original rule “if a card has a vowel on one side then it has an even number on the other”, Kirby suggests that subjects correctly infer that p and not-q instances should not occur, however, they tend not to choose the not-q card because they don’t think it is likely to reveal a p. This is so because the set of possible p cases (i.e. the vowels) is smaller than the set of possible not-p cases (i.e. the consonants) by a ratio of five to twenty-one\(^1\). It is thus much more likely that a not-p instance (i.e. a consonant) will be on the back of a not-q card, a combination that would not be informative as to the truth of the rule, so the expected utility of this card is low.
I will discuss two difficulties with this account. The first difficulty involves the choice stage. The account predicts greatly improved performance if the task were presented with the rule, "if a card has a consonant on one side then it has an even number on the other". The larger set of p instances (consonants this time) would make p more likely to occur on the back of a not-q card, and thus, Kirby would argue, many more subjects would select not-q. While no data exist for such a rule, it just seems doubtful that it would lead to such a dramatic improvement in performance. Also, Evans and Over (1996) have pointed out that likelihood of a particular instance may be based on frequency of use, in which case the p-set (i.e., the vowels) for the original rule is not as small as Kirby suggests.

The second difficulty that I have with this account involves the proposed inference stage. It seems to allow that virtually all subjects correctly infer that the p and not-q combination should not occur. The utility of this combination is presumably high, and it is only the limited likelihood of that combination which explains why not-q is not selected at the choice stage. This seems, however, like a very strong assumption. Some pilot work that I carried out with Keith Stanovich suggests, that with the abstract task, it is quite difficult to get subjects to think about which combination should not occur. We gave subjects the standard selection task rule and showed them the standard array of cards, however, instead of instructing them to find out whether the rule was true or false, we told them that they knew the rule to be false and that their task was to demonstrate that the rule was false. With this instruction, subjects could infer the likelihood of a falsifier to be 100%. If a particular falsifying combination had a high utility, then they should select cards that could reveal that combination, because under such circumstances
such a card most likely \textit{would} reveal that falsifying combination. So, if as Kirby suggests, subjects infer that p and not-q is inconsistent with the rule, they ought to have selected not-q in this version, because the likelihood of finding a p on the back is very high. We found, however, that subjects did not tend to select the not-q card any more frequently than they did so when they were given standard instructions. This suggests that it is not very likely that most subjects were even considering the p and not-q combination in the first place.

The Sperber, Cara, and Girotto (1995) account handles this result quite nicely. Since most subjects spontaneously take the rule to mean that there are p and q instances, telling them that the rule is false will only make them expect not to find any p and q instances. This would not change their pattern of selection since such instances could only be found on the p card and on the q card. In choosing those cards, subjects would be attempting to demonstrate that the hypothesized p and q instances don’t occur.

Another difficulty with Kirby’s account of the standard selection task is that he does not really offer much of an explanation as to why so many subjects incorrectly select the q card. He points out that the q card is a falsifier only if the rule is taken to be a biconditional. Johnson-Laird (1983; Johnson-Laird & Byrne, 1991) has argued for some time that q selection is the result of a biconditional interpretation of the conditional rule. Under Kirby’s account, subjects would be selecting q because at the inference stage, they infer that a not-p and q combination is inconsistent with the rule, and at the choice stage, they select q because the valuable not-p and q combination has a good likelihood of occurring. Stanovich and I, however, have evidence from another pilot study suggesting that subjects selecting the q card do not necessarily make a biconditional interpretation of
the rule. We used a procedure that gave us a way of assessing the subject’s interpretation of the rule that was independent of their final selection. Subjects were shown a selection task rule and the standard four card array. For each card they were asked of the implications for the truth of the rule if a particular feature appeared on the back. So, for example, for the p card they were first asked whether a q on the back would tell them anything about the rule. They were then asked whether a not-q on the back would tell them anything about the rule. This procedure was repeated for each card. We called the procedure the consequentialization procedure, because it forced subjects to think about the consequences of each possible card combination.

Of particular interest was the subjects’ response to the q card combinations. If a subject says that a not-p on the back of the q card would indicate that the rule was false, then, we argued, that person must hold a biconditional interpretation of the rule. If, on the other hand, they did not say that a not-p on the back of the q card would not tell them that the rule is false, then we attributed to them a conditional interpretation of the rule.

We presented subjects with either a standard abstract rule or with a version of the transport problem (Manktelow and Evans, 1979). The modal response for these rules has generally been p and q (Stanovich and West, 1998). Thirteen subjects were presented with a standard abstract rule, “If a card has an A on the letter side then it has a 6 on the number side”. Eleven of them said that not-p on the back of the q card would make the rule false, suggesting that they held a biconditional interpretation of the rule. Seven of these 11 went on to select the p and q cards when given the standard selection instruction. This result is consistent with the original interpretation of p and q selection originally
suggested by Johnson-Laird (1983), that subjects selecting p and q do so because they interpret the rule as a biconditional. This is the account upon which Kirby’s theory relies.

The results were quite different, however, when we used the transport problem. The rule that we used was “If I go to Ottawa then I take the train”. Subjects were told that the four cards were travel tickets that showed a mode of travel on one side and a destination on the other side. The p, not-p, q, and not-q cards corresponded, respectively, to “Ottawa”, “Montreal”, “train”, and “bus”. Again the critical test was the subjects’ response when asked what not-p (Montreal) on the back of the q (train) card would tell them about the truth of the rule. Only if they said that this would tell them that the rule was false, could they be interpreting the rule as a biconditional. What we found was that not a single subject (out of 12) suggested that Montreal on the back of train would make the rule false. None of the subjects could thus have been interpreting the rule as a biconditional. Yet 5 of them still selected the p and q cards. This strongly suggests that a biconditional interpretation does not necessarily underlie the majority of p and q selections, as many authors, including Kirby, assume.

Without appealing to this standard explanation of q selection, Kirby is left unable to explain this important part of selection task behaviour. Sperber and colleagues, however, do offer an account that deals with q selection while allowing that subjects make the correct conditional interpretation of the rule. Recall that under their account many subjects spontaneously infer that p and q combinations should occur. In selecting cards, they do not seek to test the conditional rule, but rather, to test their expectation that p and q combinations will occur. Such a combination could, of course, only occur on the p card or on the q card, so these are the cards that they select. Notice that this account
appeals to the notion of “expectation”. On the surface this seems similar to Kirby’s notion of “likelihood”, however, there is an important difference. The term “expectation” here also implies some kind of focusing. When a subject expects a certain outcome, she is not considering the possibility of other outcomes. In a sense, the expectation is an all-or-none phenomenon. This cannot be said of Kirby’s “likelihood”. Under that account, the likelihoods of many outcomes are compared, and this comparison contributes to the choice process. *Unlikely* combinations are considered. This is not true of *unexpected* outcomes. Only the *expected* outcome is considered. The outcomes that are not expected are not considered. For Kirby, the subject makes several inferences of the kind of outcomes that should and should not occur and then evaluates each card for its probability of revealing the more important of these outcomes. For Sperber, Cara, and Girotto (1995), there is really only one inference, and this leads to the expectation of one combination or outcome. Only the single inferred outcome is searched for.

The Sperber, Cara, and Girotto (1995) account of spontaneous inference is not a very generous view of human cognition:

Subjects trust their intuitions, that is, the output of their spontaneous inferential abilities; without any further examination, they take the directly testable consequences that they have inferred in comprehending the rule to be the consequences through which the rule should be tested. (p.53)

They argue, however, that these spontaneous inferences do lead to optimal behaviour in many situations, including most situations involving deontic statements. Also, for many non-deontic statements the pragmatics of the situation will invoke appropriate expectations, as we saw in their Prince variation of the selection task, where
subjects seemed to feel that the Prince was exaggerating his claim, and inferred that counter-examples would exist. And even in situations where spontaneous inferences mislead us, Sperber and colleagues allow for higher level “meta-inferences” to guide our behaviour. These authors provide minimal elaboration of the meta-inferential process, however, the idea does deserve some attention.

**Compatibility with dual processing accounts**

The notion of meta-inference is compatible with aspects of the dual-processing accounts proposed by Evans and Over (1996), and by Sloman (1996). While Evans and Over call the respective systems “heuristic” and “analytic”, Sloman calls them “associative” and “rule-based”. The associative system is fast and relies on cues of similarity. The rule-based system operates on symbolic representations and could thus be better equipped to analyze the truth of a given statement. Viewed like this, the spontaneous inferences proposed by Sperber, Cara, and Girotto (1995) could be the product of Sloman’s associative processes. Such processes could guide comprehension of ordinary conversation, and may thus also guide people’s interpretation of reasoning tasks like Wason’s selection task. The meta-inferences, which Sperber and colleagues suggest are rarely made by subjects performing the selection task, could be the product of Sloman’s rule-based system. Sperber and colleagues describe a meta-inference as an inference resulting from a more formal method of evaluating the truth of a statement.

In his discussion of dual processing accounts, Stanovich (in press) suggests that the rule-based system may “override” (following Pollock, 1995) the associative system in certain situations. This will happen if the subjects is motivated to ignore or question the output of the associative system, so long as the individual has the time and capacity to do so. This is compatible with the relationship of meta-inference to spontaneous inference
in the Sperber and colleagues account. Meta-inference could take over if the spontaneous inferences that the subject generates do not seem sufficiently relevant. Recall that spontaneous inferences are made until relevance is achieved. But it may be that the typical spontaneous inferences made in the standard selection task (i.e., “p and q instances should occur”) may not seem satisfactory to the subject. The subject may infer that the experimenter must intend something more complex than that. Such test savvy may prompt a new search for an appropriate interpretation, this time, using a more methodical process involving consideration of all possibilities. This may lead to correct selection if it leads the subject to consider each card and to recognize that the not-q card could reveal an instance of “p and not-q” which would make the rule false.

It should be noted that meta-inferences may not be successful. For some subjects, meta-inference may fail to produce any relevant insight. Such subjects may indeed give consideration to all of the cards, but, perhaps due to individual differences in ability, or to a temporary lapse in concentration, such subjects may not correctly deduce the consequences of turning over the different cards. In such situations they may resort to their previous spontaneous inferences.

Even successful meta-inference needn’t always lead to the optimal p and not-q response. For subjects selecting the p card only, meta-inference may be involved as follows. The subject makes the spontaneous inference that p and q instances should exist. In considering the p card they notice that if this card does not reveal a p and q instance, then it must reveal a p and not-q instance which would make the rule false even if the q card revealed a p and q instance. They also notice that if the p card does reveal a q, then the spontaneously inferred consequence will have been satisfied, and revealing another p
and q instance, this time on the q card, would not be informative. In either case, selecting the q card would be superfluous.

The p-only selection has not generally been regarded as insightful in the selection task literature. The above analysis suggests that such a selection, while inadequate in its lack of recognition of the not-q card, may be very insightful in its non-trivial recognition that the q card does not have any value in evaluating the truth of the conditional statement. Stanovich and West (1998) have recently provided evidence that subjects who make this selection may be more capable than most subjects when it comes to what we might call meta-inference. This suggests that a high level insight may thus underlie this selection.

Stanovich and West used this kind of argument to suggest that p and not-q was in fact that normative selection in the selection task. This had long been taken for granted in the literature but had recently come under attack by several authors (Margolis, 1987; Oaksford & Chater, 1994). Stanovich and West used Anderson’s (1990, 1991) ecological model to argue that the selection made by the most intelligent individuals is likely to be the one that is normative. Such individuals are the ones with the most adaptive thought processes, and hence, it could be argued that the output of these thought processes is most likely to be optimal. The possibility that the small minority of subjects who make the p and not-q selection was of higher average intelligence had generally been ignored in the literature. It had been assumed that such subjects represented little more than random noise. Stanovich and West showed that such subjects had a higher average SAT score than did any other selection group. SAT has been shown to correlate with various indicators of intelligence, including working memory capacity.
An unexpected finding in the Stanovich and West study was that the group of subjects who selected p only on the standard selection task also had high SAT scores. Their average was higher than that of the p and q pickers, but also higher than the subjects who selected p, q and not-q. This latter selection has traditionally been called the “partial insight” selection (Wason & Johnson-Laird, 1972), but the Stanovich and West result suggests that the subjects that make this selection are actually less likely to be the ones able to make an insightful selection. The p only selection may thus be better considered the “partial insight” selection. The partial insight is thus the recognition that q needn’t be selected, rather than the recognition that not-q needs to be selected. In fact, across all subjects, those who did not select q had higher SAT scores than those who did select not-q. This suggests that more cognitive resources are required to make the former recognition than the latter.

**Conclusion**

In this study, the Relevance account of selection task performance (Sperber, Cara & Girotto, 1995) has been tested against competing accounts in two experiments. In the first experiment it was tested against Cummins (1996a) modularity account which predicted that expectation effects would not affect reasoning behaviour in young children, and that performance would be exclusively determined by the deontic/indicative nature of the presented scenario. With the small sample size, a statistically significant result was not observed, however, there was a trend suggesting that an indicative scenario, where children expected violation, was more likely to lead to violation detection behaviour than was an indicative situation where violation was not expected. This trend is consistent with the Relevance account, which suggests that people will make inferences depending
on the pragmatics of the scenario, and that these inferences will be in the form of expectations. It is these expectations that are thought to underlie selection behaviour.

In Experiment 2, it was found that with the abstract selection task, selections could not be based solely on the subject’s having formed mental models of particular cards and cases, as suggested by the theory of mental models (Johnson-Laird & Byrne, 1991). It was demonstrated that forming a mental model of the falsifying combination does not necessarily lead to selection of the not-q card. This was seen when a group of subjects in the falsification and confirmation condition was less likely to select the not-q card than was a group in the falsification only condition, even though both groups had showed evidence of having successfully represented the p and not-q combination as an outcome that should not occur. It seems that in addition to forming a mental model of the card combination, a subject must also feel that such outcomes are either likely or important.

It has been argued that the results of both experiments can be explained by the Relevance account offered by Sperber, Cara, and Girotto (1995). While two other theories -- the dual processing theory of Evans and Over (1996) and the subjective expected utility theory of Kirby (1994) -- may also explain the present results, it was argued that the Relevance theory is to be preferred on both empirical and theoretical grounds. The Relevance account is, however, consistent with aspects of the dual process account offered by Sloman (1996). It was suggested that what Sperber and colleagues call “spontaneous inferences” could be the product of Sloman’s fast “associative system”, while their “meta-inferences” could be the product of the more methodical “rule-based system”.
The nature of these different kinds of inferences needs to be explored more carefully. Consideration of Bartsch and Wellman's (1995) suggestion that young children can more readily think of other people's desires than of their beliefs, could lead to investigation of the possibility that spontaneous inferences are more readily influenced by the desires of others than by their beliefs. By having children report on the desires and beliefs of different characters, it may be possible to observe differences in how well each predicts selection behaviour in tasks like that of Experiment 1.

With respect to meta-inferences, Sperber and colleagues argue that the selection task is not the right tool for exploring the nature of this process, because the vast majority of subjects make selections based on spontaneous inference. Stanovich and West (1998), however, have argued that there is evidence of some kind of meta-inference for some subjects, and it may be possible to manipulate the instructions to increase the use of meta-inference. One possibility would be to encourage subjects to think through all of their options before selecting. This was the basis of the consequentialization pilot study conducted by Stanovich and myself (described earlier). Alternatively, telling subjects that their initial response to a standard selection task is "almost certainly wrong" may prompt a more methodical approach to the problem and produce insightful meta-inferences. This instruction may work quite differently from offering monetary incentives for correct responses. The latter approach will not necessarily make a subject question her answer. What is needed is for the subject to mistrust her initial spontaneous inference. The mistrust may lead the rule-based system to "override" (to use Pollock's, 1995 terminology), the associative system, and thus produce meta-inferences. Evidence of such an override may be present in subjects' verbal protocols. If the meta-inferential
process is indeed slower, a subject's rate of speech may slow down, or she make be heard to take the time to consider the implications of different cards.

It may, however, turn out that such an override is very difficult to invoke. In many situations, it is often difficult for people to escape their initial intuitions. Turning to rule-based reasoning and producing meta-inferences may simply not be something that people can readily do. Stanovich and West's (in press) observed individual differences in use of a meta-inferential strategy may reflect a significant capacity or tendency that only a minority of people develop. If this is indeed the case, then close study of such individuals would be warranted in the hope of understanding how the ability develops. While spontaneous inference remains an important and useful aspect of human cognition, meta-inferential abilities are becoming ever more important in the modern, literate and technological society where one's life is affected by events in all corners of the world. Consequences of actions and events need to be more closely examined than ever, and only comprehensive meta-inferences allow that kind of analysis.
Note

1. Oaksford and Chater (1994) have shown that considerations of card frequency could explain performance on very many versions of the selection task. They argue that in most situations, for a conditional rule of the form "if p then q", p instances will be much less common than not-p instances, and q instances will be much less common than not-q instances. They call this the "rarity" condition since instances of p and of q will typically be rare in the environment. Under such a condition, looking at instances of not-q is not likely to be helpful, because there are so many not-p instances compared to p instances in the environment, that even if p and not-q cases do exist, there will be many more not-p and not-q cases. The most efficient search in such an environment is to seek confirming instances of p and q, the presence of which will increases the likelihood that the statement is true. After arguing that such a strategy is optimal in the environment in which humans most commonly find themselves, Oaksford and Chater suggest that subjects assume that the rarity condition holds for the standard Wason selection task, and thus the modal p and q selection is in fact the normative response.
References


Table 1: Percentage of correct "Yard" selections in Experiment 1.

<table>
<thead>
<tr>
<th></th>
<th>Original Deontic n=17</th>
<th>Original Indicative n=18</th>
<th>Modified Indicative n=18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage</td>
<td>71</td>
<td>33</td>
<td>56</td>
</tr>
</tbody>
</table>
Table 2: Number of subjects in each condition correctly performing the several stages of Experiment 2.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Falsification Only</th>
<th>Falsification and Confirmation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Envisaged P and ( -Q ) as the falsifying combination</td>
<td>69</td>
<td>72</td>
</tr>
<tr>
<td>Envisaged correctly and identified P and ( -Q ) as the only cards on which the falsifying combination could occur</td>
<td>36</td>
<td>33</td>
</tr>
<tr>
<td>Envisaged and identified P and ( -Q ) for the falsifying combination, and envisaged and identified P and Q for the confirming combination</td>
<td>N/A</td>
<td>21</td>
</tr>
<tr>
<td>Correctly envisaged and identified the falsifying combination and (if required) envisaged and identified P and Q as the confirming condition, and went off to select ( -Q ) during selection.</td>
<td>21 ( (58%) )</td>
<td>5 ( (24%) )</td>
</tr>
</tbody>
</table>
Table 3: Number of subjects making particular selections in the Falsification only and the Falsification and Confirmation conditions for Experiment 2.

<table>
<thead>
<tr>
<th>Selection</th>
<th>Falsification Only N=93</th>
<th>Falsification and Confirmation N=95</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>-P</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Q</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>-Q</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P and Q</td>
<td>27</td>
<td>33</td>
</tr>
<tr>
<td>-P and Q</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>P and -Q</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td>-P and -Q</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P, Q and -Q</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>None</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>All</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>Other</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>
Table 4: Number of subjects in each condition envisaging P & -Q, -P & Q, both, or “other” as the combination that could falsify the rule in Experiment 2.

<table>
<thead>
<tr>
<th>Falsifying Combination</th>
<th>Falsification Only n=93</th>
<th>Falsification and Confirmation n=95</th>
</tr>
</thead>
<tbody>
<tr>
<td>P &amp; -Q</td>
<td>69</td>
<td>72</td>
</tr>
<tr>
<td>-P &amp; Q</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Both P &amp; -Q and -P &amp; Q</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>10</td>
<td>15</td>
</tr>
</tbody>
</table>
Table 5: Number of subjects in the Falsification and Confirmation condition envisaging P & Q, -P & -Q, both, or “other” as the combination that could occur if the rule were true in Experiment 2.

<table>
<thead>
<tr>
<th>Confirming Combination</th>
<th>Number of Subjects Envisaging (Falsification and Confirmation condition only)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=95</td>
</tr>
<tr>
<td>P &amp; Q</td>
<td>73</td>
</tr>
<tr>
<td>-P &amp; -Q</td>
<td>8</td>
</tr>
<tr>
<td>Both P &amp; Q And -P &amp; -Q</td>
<td>3</td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
</tr>
</tbody>
</table>
Table 6: Number of subjects identifying given combinations as potential falsifying instances in the Falsification only condition. (The bracketed number indicates the subjects who selected exactly the cards that they identified as potentially revealing a falsifying instance).

Subjects envisaging P & -Q as the falsifying combination (n=60)

<table>
<thead>
<tr>
<th>Card(s) identified as being a possible falsifying instance</th>
<th>Card(s) identified as being a possible falsifying instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>P only</td>
<td>-P only</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>(2)</td>
<td>(0)</td>
</tr>
</tbody>
</table>
Table 7: Number of subjects identifying given combinations as potential falsifying and confirming instances. (The bracketed number indicates the subjects who selected exactly the cards that they identified as potentially revealing either a confirming or falsifying instance).

Subjects envisaging P & -Q as the falsifying combination (n=53)  

<table>
<thead>
<tr>
<th>Card(s) identified as being a possible falsifying instance</th>
<th>Card(s) identified as being a possible falsifying instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>P only</td>
<td>-P only</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>P only</td>
<td>5</td>
</tr>
<tr>
<td>-P only</td>
<td>0</td>
</tr>
<tr>
<td>Q only</td>
<td>4</td>
</tr>
<tr>
<td>-Q only</td>
<td>0</td>
</tr>
<tr>
<td>P Q</td>
<td>2</td>
</tr>
<tr>
<td>-P Q</td>
<td>0</td>
</tr>
<tr>
<td>P -Q</td>
<td>0</td>
</tr>
<tr>
<td>-P -Q</td>
<td>0</td>
</tr>
<tr>
<td>P Q -Q</td>
<td>0</td>
</tr>
<tr>
<td>all</td>
<td>0</td>
</tr>
<tr>
<td>none</td>
<td>0</td>
</tr>
<tr>
<td>other</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 8: Number of subjects identifying given combinations as potential falsifying and confirming instances. (The bracketed number indicates the subjects who selected exactly the cards that they identified as potentially revealing either a confirming or falsifying instance).

Subjects envisaging $P \& -Q$ as the falsifying combination ($n=11$)

Subjects envisaging $-P \& Q$ as the falsifying combination ($n=0$).

<table>
<thead>
<tr>
<th>Card(s) identified as being a possible falsifying instance</th>
<th>Card(s) identified as being a possible falsifying instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>P only</td>
<td>-P only</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>P only</td>
<td></td>
</tr>
<tr>
<td>-P only</td>
<td></td>
</tr>
<tr>
<td>Q only</td>
<td></td>
</tr>
<tr>
<td>-Q only</td>
<td></td>
</tr>
<tr>
<td>P Q</td>
<td></td>
</tr>
<tr>
<td>-P Q</td>
<td></td>
</tr>
<tr>
<td>P -Q</td>
<td></td>
</tr>
<tr>
<td>-P -Q</td>
<td></td>
</tr>
<tr>
<td>P Q -Q</td>
<td></td>
</tr>
<tr>
<td>all</td>
<td></td>
</tr>
<tr>
<td>none</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td></td>
</tr>
</tbody>
</table>
Appendix A:

Other Results from Experiment 2

The final selections of the subjects in both conditions of Experiment 2 are shown in Table 3. As expected, more subjects selected p and not-q in the *falsification only* condition. The only other major difference in the pattern of selection was that more subjects in the *falsification and confirmation* condition selected p, q, and not-q than did so in the *falsification only* condition. This is not surprising since the former group were more likely to expect instances of p and q because of the task instruction, and hence, would be more likely to select the q card.

The combinations that subjects envisaged as the falsifier were similar across conditions. This is shown in Table 4. This result was expected since the falsification task was performed first in the *falsification and confirmation* condition, so the presence of the confirmation task should not have been a confound. In both conditions the majority of subjects envisaged the p and not-q combination as the falsifier (74% and 76% respectively). A small percentage of subjects in both conditions envisaged not-p and q as the falsifying combination, or envisaged both not-p & q, and p & not-q (15% in *falsification only* and 8% in *falsification and confirmation*). These responses are appropriate if the subject had made a biconditional interpretation of the rule. The envisaging task does not permit much by way of analysis of rule interpretation, however, as the modal response of p and not-q could also be given by a subject making a biconditional interpretation. The figures of 15% and 8% may then be underestimates of the proportion of subjects making such an interpretation.

Table 5 shows the combinations given as a confirming combination by subjects doing the second envisaging task in the *falsification and confirmation* condition. Seventy-seven percent of subjects offered p and q as the confirming combination. Twelve percent
envisaged either not-p and not-q or both combinations. The other twelve percent of subjects did not suggest a meaningful combination, indicating that they probably did not understand part of the task.

Tables 6, 7, and 8 show the cards that were identified by subjects after they envisaged particular combinations. The tables also show the numbers of subjects who selected exactly the cards that they had identified. There was relatively little clear clustering of responses, so no analysis of this data is offered beyond that which is discussed in the "Results and Discussion" section.