# Using Novel Semantic and Informational Manipulations of Rules to Extend an Interpretive Approach to Conditional Reasoning

*Eshaa Mohamed Alkhalifa*

Doctor of Philosophy
Institute of Communicating and Collaborative Systems
School of Informatics
University of Edinburgh
2004

# Abstract

This thesis extends a view of human reasoning which emphasises a theory of interpretation in conditional reasoning. It extends work on Wason's (1968) 'selection task', using novel rules and contexts to explore the factors that control subject interpretations, which in turn is reflected in their performance. After reviewing the work on conditional reasoning and particularly the interpretative framework of Stenning & van Lambalgen (2004), the thesis explores subjects' reasoning with rules that describe processes extended in time in two experiments. The most striking finding is that many subjects exhibit an unusual constant anaphor reading, even though the anaphors involved are tenses rather than pronouns. Results are explained in terms of the temporal constraints involved in the situation described.

The thesis then uses novel 'information packaging' manipulations which use colour to emphasise different distinctions in Wason's original task. This manipulation provides evidence of where subjects' attention already rests. This is combined with a task that gathers data of subjects' interpretation of negation. Results are consistent with the idea that although subjects in the standard task are focussed on the distinction between cases that fit the rule and ones that do not, there is evidence that emphasising the mapping of the antecedent/consequent onto back/front of the cards is sensitive to these manipulations. The negation interpretation task reveals striking divergences between subjects' interpretations and the classical model assumed in the literature, and these differences are interpretable in terms of default logic. A few conditions were originally designed as controls only to end up generating striking results of their own. Colour is used in the truth conditional semantics of the rules (black/white replaces number/letter or vowel/consonant) instead of being used as mere information packaging. Sizable increases in 'classical competence' responses are observed and this is interpreted in terms of the non-hierarchical structure of the properties used. Studies using LSA and a novel tensor network operating on a database of rules gathered from selection task literature show conclusively that higher frequencies of function words appear in descriptive rules than they do in deontic rules.

This thesis concludes that it is possible to direct subjects towards various interpretations on the task through the use of semantic manipulations that include but are not restricted to the ones observed in this work. Issues that include resolving the anaphora in the problem, the hierarchy of the structure of properties and the negation of clauses clearly influence the interpretation subjects arrive at which in turn affects their reasoning and responses. Results indicate that subject assumptions concerning negatives are different from those made by experimenters which inform all major theories including Mental Models, Information Gain and those based on mental logic. In particular, subjects' most frequent selections indicate that they are selecting an implicit negative which is the opposite of what is expected by the principle of truth of Mental Models.

# Acknowledgement

First and foremost, one may only bow with respect to a prophet of true greatness as to instil a love for knowledge as Mohamed Ibn Abdulla may peace be upon him.

Sincere thanks and appreciation goes to His highness Shaikh Isa bin Salman Alkhalifa the father who encouraged all to learn and who financially supported this study.

I would also like to thank my husband Nawaf E. Alkhalifa, for enduring a long process, continual travel, and horrendous expenses during the period of study.

Special thanks go to my sons, Ebrahim and Mohamed as only through their patience and encouragement, I felt no guilt for being away from them so much of the time.

I would also like to thank my principal supervisor Prof. Keith Stenning, who is an undisputed expert in helping one realise that there is more to knowledge than believing that you know what you found.

Last but not least, I would like to thank every person who gave me a word of support, and every person who wrote a word that added to my knowledge.  You are too many to list here but I appreciated every individual who helped make the work here arrive to this stage.

# Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(Eshaa M. Alkhalifa)

# Dedication

*To God*

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

> *"Pure mathematics is the subject in which we can be perfectly certain of the truth of what we say at the expense of not knowing what we are talking about" (Bertrand Russell)*

A central concern of the psychology of reasoning is to ascertain to what extent a person can draw conclusions that follow through logic alone if given a set of assumptions (Wason and Johnson-Laird, 1972). The crucial point that commands your attention is that the claim is that subjects start their reasoning given a set of assumptions whereas in fact they are reasoning from a textual question from which they are expected to extract their set of assumptions.

This thesis is concerned with analysing the relation between the interpretations that subjects make of the textual question and their reasoning which is based upon these interpretations in addition to the way the two interact. Several researchers have recognised in the context of the selection task as an example (Manktelow and Over, 1978; Stenning and van Lambalgen, 1999) that subject interpretations of textual questions diverge from what is expected by classical models. Natural language texts given in the questions may offer multiple possible semantic implications if subjects assume different contexts or make erroneous assumptions. By contrast, classical logic models presume that the same set of assumptions will always be made independent of the semantic implications made. Evidently, this is not the case.

Stenning and van Lambalgen (2004) offered an interpretive account where subjects impose interpretations onto the task, and semantic concepts such as anaphora and contingency can help explain their responses. This thesis therefore, takes the selection task as a representative as was done by Stenning and van Lambalgen (2004) and extends the view they proposed.

The experiments exhibited here impose different semantic structures onto the task to yield different response distributions among possible choices. None of the existing theories offer any explanation of subject behaviour in the task, as the manipulations are semantic as those in the work of Stenning and van Lambalgen (2004). The semantic structures imposed onto the task include the use of rules about processes that imply a temporal order between the events described in the antecedent and the event described in the consequent. "Information packaging" was also utilised as a tool of analysis to identify the role that the background rule plays in reasoning and where subjects place distinctions to group the information they are given. This thesis also challenges a basic assumption by questioning what subjects regard as negatives of each other.

The main claim is that novel manipulations of content influence subject distributions among the different responses possible. The effect-response pattern is not at all trivial but it fills a part of the puzzle of how interpretation and reasoning interact. This in turn, has broad implications for reasoning and cognition in general.

The thesis will start with a review of prior work and then move on to the interpretive perspective offered by Stenning and van Lambalgen (2004). It builds on that by introducing manipulations that impose temporal sequencing and follows that with information packaging to impose semantic distinctions onto the task. Then it analyses the surface structure of the rules, through a tensor based neural network.

## 1.1 Background

The quest for the target defined by Wason and Johnson-Laird (1972) has diversified into many paths of research depending on the type of tasks chosen as a medium of analysis. These tasks include, amongst others, construction tasks, evaluation tasks, as well as the selection task. Construction tasks provide a subject with an incomplete sentence like "… *is an even number*" and the subject is asked to complete the sentence such that it is true or false. For example, if asked to answer it such that it is true a possible answer will be "*2 is an even number*". Evaluation tasks on the other hand, present a subject with a completed sentence in a context, and ask subjects to indicate whether the sentence is true or false in that context. The selection task, similar to both, yet unlike them, presents subjects with a rule, and four possible test cases from which they must select the minimum number they can in order to ensure that the rule holds. The difference that is clearly evident is that the Construction task asks subjects to insert a value from a range that has two sets, while the Evaluation task asks subjects to evaluate the sentence to identify if the two values it contains are aligned. On the other hand, the selection task asks subjects to evaluate a relationship between two sets and whether or not it exists within a four card domain. This is clearly a more complex task when compared to the other two, due to the larger range of possibilities and misunderstandings that may arise. It involves verification and falsification which are two mental processes that are still ill understood.

The selection task as proposed by Wason (1966) has results that seemed shocking at first sight. Four cards are presented to students who were told that the cards had a number on one side and a letter on the other. The cards show an "*A*", a "*B*", a "*4*" and a "*7*". They are told to check whether or not a rule holds: If a vowel appears on one side then an even number is on the other side of the card. Subjects have to show which cards they wish to turn over to check the rule. The correct responses according to classical logic would thus be A and 7, or in other words to choose the true antecedent and the false consequent cards. Oddly enough, only between 4% and just over 20% of students selected these two cards in empirical studies (Manktelow & Evans, 1979). This is odd because the given rule is a straightforward *If P then Q* rule of the sort that is understood even by young children that are told of rewards for specific behaviour or punishment if they misbehave. It is quite common for a child to be told: "*If you finish your homework early, then you can go play on your bike*" while assuming that the child will understand the rule. Yet, it is the process of validating this extremely common rule that such a large number of subjects get wrong. This result started nothing short of a gold rush with the aims of finding the magical explanation to this behaviour.

Since subjects' observed behaviour was regarded as an "*error*", initial work was determined to find cases that would not lead to this "*error*". This resulted in several theories that basically divided materials into two groupings; abstract versus "*thematic*" materials.

"Thematic" materials were observed to result in a much higher level of accuracy (Wason & Shapiro 1971, Wason & Johnson-Laird 1972). These materials were characterised with a "thematic" relationship between the two propositions in the conditional. An example of this type of problem is: "*If I go to Manchester, I go by train*"[1] (Wason & Shapiro, 1971). On the other hand, 'abstract' materials included

[1] This particular example is an elusive case that failed to replicate and only worked for Wason and Shapiro (1971), however it will be discussed later.

materials similar to the example given above and were usually characterised by being ‘concrete’ but lacking in any “thematic” linkage between the propositions.

The assumption is that while abstract materials do not seem to guide subjects to presume that a negative case may exist, “thematic” materials do imply possible negative cases that have to be checked. Since classical logic implies that the correct selections to check a rule of the form *If P then Q* are to select P and NotQ, then a possible false instance would make subjects aware that NotQ must be checked as well. However, the dividing line was never clear-cut and almost no linguistic or logical insight was brought to illuminate the possible causes of the poor performance in the abstract tasks that are also described as arbitrary tasks.

> *“Some of these patterns are inconsistent with any logical interpretation of the materials.”(Cheng & Holyoak, 1985)*

Unfortunately, researchers quickly jumped to the conclusion that results could not be explained through any logical interpretation and were satisfied with the division of materials into these two groups claiming that the premise ‘semantics’ played a role by raising the possibility of a falsifying instance.

> *“The model theory predicts that any manipulation that emphasizes what would falsify the rule should improve performance in the selection task.” (Johnson-Laird, 1999)*

Therefore it seems to be a safe assumption to make that according to these conclusions, the abstract task questions would result in ‘inaccurate’ reasoning. They would lead subjects astray to a diverse set of interpretations while “thematic” type rules would lead subjects to seek falsifying instances that need to be checked to ensure that the rule is correct, thereby achieving the classical logic response to the question.

## 1.2 A Theory in Logic

> *“.. form will interact with content in reasoning, and the goal of the theory is to understand how” (Stenning, van Lambalgen, 1999)*

Stenning and van Lambalgen (1999) indicate that one of the main neglected issues in prior work is that the form of the rule is not fully understood with respect to how it interacts with the content or the semantics of the task. The form is what dictates the “*expected*” answers with respect to classical logic or Popper’s (1963) theory, while content is the specifics of the "*thematic*" relations that influences performance in the task. In order to investigate this issue further, Stenning and Lambalgen (2002b) introduce a new conceptual perspective to this task by showing that logical theory expects the form subjects impose on the task is strongly affected by the semantics of the task itself. This means that subjects do not have a unified understanding of the task form. Instead, they have one of several different possible "*understandings*" and the one they select influences how they approach the task and what selections are the "*correct*" ones to make.

However, since the "*expected*" answers are extracted from a classical view of logic, it does take into account that the logic subjects use may be using is semantically sensitive such as deontic or real world logic. Stenning and van Lambalgen (2002b) show through Socratic dialogue with subjects that indeed interpretations differ greatly from each other with respect to how that particular subject understood the form of the abstract task.

This finding explains the distinction found between the two types of tasks, namely, abstract and "thematic", and allowed Stenning and van Lambalgen (2002b) to redefine them as descriptive and deontic tasks. In a descriptive question the rule can be: "*If there is an A on one side, then there is a 4 on the other side*". If a case is found where there is an A on one side and a 7 on the other, then it is natural to assume that this rule is false at least for this case. On the other hand, if the rule is: "*If a customer is to drink an alcoholic beverage, then she must be over 21"* and someone under that age drinks alcohol, then the case would be considered a violation of the rule. Its existence does not pose any threat to the existence or truth of the rule, making this a deontic rule, while the existence of the first case threatens the truth of that rule, making it a descriptive rule. This clarifies the distinction between successful cases and unsuccessful ones which are strongly dependent on whether or not the rule is fragile or reliably resistant to the existence of a falsifying instance.

Conversely, descriptive tasks do not lead subjects to follow the path of correct deontic reasoning, thereby causing them to arrive at a diverse medley of possible conclusions depending on the types of interpretations they adopt. Stenning and Lambalgen (submitted) outline several different possible interpretations that subjects may arrive at in the case of the descriptive task that in the simplest terms makes that version confusing. These include interpreting the task such that "*not false*" is not interpreted the same as "*true*". Another interpretation subjects assume is when they negate the conditional of the form "*If P then Q*" as "*If P then Not Q*". A third interpretation that is possible is that each card has its own domain versus the four being part of the same domain, while the fourth interpretation is that some rules have exceptions. Other issues include taking the cards as a sample of all letters and numbers and dependency issues between card choices, such as if one card is turned and gets some result how this would affect future choices of cards.

However, there still remain many unresolved issues such as why subjects make a higher selection percentage of PQ as preferred to other popular subject choices such as only P and only Q. Nor is there strong experimental evidence that shows through a probe into subject interpretations that they do indeed have different interpretations of what NotP and NotQ are indicating and that they did have a diverse set of interpretations of the form of the rule. The theory has not as yet investigated the relationship between the foreground rule and the background rule, nor the semantic implications of the word "*not*" versus giving another possible value as in comparing A, Not A to A, B. This would be especially interesting if the "*not*" is expressed semantically so as not to strongly affect the form of the comparison. These are the issues that will be under consideration in this thesis.

## 1.3 Semantic Games

> *"Two men went camping overnight in the desert and slept peacefully through part of the night. They woke up, and one asked the other: What do you see? The other answered: A dark sky full of beautiful stars. The first asked again: What can you deduce from that? The other answered: ..that we are only a tiny speck in a vast universe. The first responded sharply: You idiot! It means that our tent has been stolen."*[2]

The task at hand therefore, requires the introduction of novel semantic tools with which subject interpretations can be directed towards a different set or different distribution. The aim is to make the experimental setting "*invite*" different interpretations. Perhaps the ideal starting point is from the division of materials that distinguishes materials that yield the P, Not Q responses versus those that do not.

Stenning and van Lambalgen (2002b) introduce Deontic logic as a 'real life' logic that explains the difference between descriptive/abstract and deontic/"thematic" questions. One of the aspects of this type of reasoning is that it compares the status quo with an

[2] There are several versions of this joke roaming the world, the most famous claims the two are Sherlock Holmes and his aide Watson.

ideal, when it discusses what "*ought*" to be done. A rule can be represented by the form O(α|β) which is *"α ought to be done if β is done"*.

A feature exists in 'deontic logics' that is especially interesting; namely, a paradox by the name of the Chisholm paradox that is represented whenever the following sequence exists (van der Torre & Tan, 1998).

**1a.** α ought to be (done)

**2a.** if α is (done), then β ought to be (done).

**3a.** If α is not (done), then β ought not to be (done). {This is understood from the context}

**4a.** α is not (done)

Now the paradox arises from mapping "***ought to be (done)***" in line 1a to "***is done***" in line 2a, and in a sense this comes from applying an "*ideal*" to a real life case and it results in a recommendation that β ought to be (done). Then by mapping 4a to 3a we also arrive at the conclusion that β ought not to be (done) which contradicts the conclusion arrived at above and consequently is paradoxical. The first mapping may raise a concern that 1a and the first premise of 2a do not match perfectly, however it should be evident that in real life when applying ideals to real life situations, this type of mapping is common. Now take a look at the same paradox in the present perfect:

**1b.** α ought to be (done)

**2b.** if α **has been** (done), then β ought to be (done).

**3b.** If α **has not been** (done), then β ought not to be (done). {This is understood from the context}

**4b.** α **has not been** (done)

Now the difference here from the above is that the mapping from 1b to 2b is different in a sense that 2b does not reflect an action that takes place in accordance with the ideal presented in 1b but instead it reflects a conditional made about an incident that took place in the past. Since 4b states that α **has not been** (done), then no contradiction arises in this case.

If the alteration of tenses can have this effect on this paradox, then it may also have an effect on subject performance. For example, if the temporal tag between the two premises always enforces a specific order that the first for example occurs before the second as with the line 3b then subject behaviour in the abstract/descriptive task could be altered from the majority selections of PQ.

Consequently, the work that is done here builds on this sensitivity to temporal distance in semantics between the present and past tenses. An experiment is presented through altering the temporal distances between the two premises to show that subjects rather than select PQ, select either the P, which comes first, or the Q which follows several minutes later. This separation is then fortified through yet another experiment that shows a conveyor rather than a verbal description of the time difference.

In order to find out other semantic tools through which we can analyse the selection task, a short visit to analytical reasoning seems in order. Stenning (2002) drew a map of how analogical reasoning operates at a higher level of abstraction than the events or objects described. He starts off by giving two similar stories that can be seen to map analogically, taken from Gentner et. al. (1993)

**Base Story**

> *Karla, an old hawk, lived at the top of a tall oak tree. One afternoon, she saw a hunter on the ground with a bow and some crude arrows that had no feathers. The hunter took aim and shot at the hawk but missed. Karla knew the hunter wanted her feathers so she glided down to the hunter and offered to give him a few. The hunter was so grateful that he pledged never to shoot at a hawk again. He went off and shot deer instead.*

**Analogy Match**

> *Once there was a small country called Zerdia that learned to make the world's smartest computer. One day Zerdia was attacked by its warlike neighbour Gagrach. But the missiles were badly aimed and the attack failed. The Zerdian government realized that Gagrach wanted Zerdian computers so it offered to sell some computers to the country. The government of Gagrach was very pleased. It promised never to attack Zerdia again.*

Now if we consider these two stories, we find their analogical similarity clear in the relationships that hold between the various players in each. Karla the hawk is in a sense similar to the small country Zerdia that had something that another warlike player desired, and the reactions were the same. Here, we do not have any one of the two expressed in a purely abstract fashion as in the case of the rule relating a letter to a number. Instead, what we have is a description of two scenes in which both stories are in context, and the essential features necessary for the analogy to hold consist of the way the different players interact with each other.

Since this interaction is of such crucial important in analogy then why should the same not apply to the selection task. Sure enough in the selection task, rules that are descriptive in nature resulted in very poor selection rates of the expected P, Not Q responses so the definition of this category was modified from the word abstract to the word descriptive because it seems to better describe it. The "*successful*" group also started with a general term which is "*thematic*" and that was replaced with "*deontic*" because a relationship that defines what "*ought to*" be the case seemed to exist in most if not all "*successful*" rules either within the rule itself or implied through the context.

This qualified the deontic group of rules to have a relationship imposed between the antecedent and consequent clauses that in turn resulted in many theories aimed at explaining behaviour in this particular group as in Pragmatic Schemas (Cheng and Holyoak, 1985;Cheng et. al.,1986) and Social Contracts (Cosmides, 1989). These theories lacked generality and, for that, other theories emerged including Relevance Theory (Sperber et al., 1994) and Information Gain (Oaksford and Chater, 1995). Yet again the work presented here does not aim to judge the two theories or to select between them.

Instead, what is done here is an investigation into the relationship between the antecedent and consequent for descriptive rules. In order to achieve that target let us take a look at a modified version of the analogy stories given above as presented by Stenning (2002).

> *Once there was a small country called Zerdia that learned to make the world's smartest computer. One day the chief minister of its feeble neighbour Gagrach challenged Zerdia to a national tournament of Trivial Pursuits. Gagrach lost badly and a gloom descended. The Zerdian government realized that Gagrach needed Zerdian computers to help its team's memory, and so it offered to sell it some. The government of Gagrach was very pleased. It promised, in the future, to only challenge other countries to Trivial Pursuits.*

Now we can consider this story in two ways; the first is to consider it appearing on its own. In this case, one may assume the word "*trivial*" in "*Trivial Pursuits*" to imply the game is a simple one that is not at all alarming in nature. One may also assume that Zerdia decided to help Gagrach out of kindness and having the same computers implies that the two cannot challenge each other to this game perhaps because of the game rules.

If, on the other hand this story is compared to the base story, several issues immediately arise. The actions that are performed and intentions map perfectly onto the base story. If then an analogy is drawn, the game of Trivial Pursuits is understood to be a game that is not to be desired in spite of what is implied by its name "*trivial*". Another issue is that the neighbour here is "*feeble*" so it is not understood to be threatening except perhaps by this mysterious challenge that it promised to never do again.

Consequently, what happened here is that the intended interpretation of the term "*Trivial Pursuits*" was altered through the analogical mapping process without changing a single word in the story. It was initially assumed to imply a straightforward interpretation and was altered to imply an ironic title. This type of change is purely semantic so we can consider applying it to the context of the selection task rule. In order to do that, we must mimic what occurred to the term with minimal changes to the rule.

The rule "*If there is an A on one side, then there is a 4 on the other side*" differs from the rule "*If there is a black A on one side, then there is a white 4 on the other side*" by virtue of the imposed semantic distinction between what is on the face of the card and what is on the back of the card. Black and white are opposite colours and here they are utilised to identify if subjects impose a distinction between the two sides of the cards by virtue of applying the background rule. Other distinctions can be probed in a similar fashion.

If subjects already emphasise a distinction that is emphasised through colour then no difference in their behaviour is expected just as the base story seems to map well to its analogical match. However, if the emphasised distinction differs from what is dictated by subject interpretations, then a difference in behaviour is expected as is expected when mapping the base story to the "*Trivial Pursuits*" story.

Consequently, the ideal means to analyse descriptive rule behaviour is through the use of "*information packaging*" in order to emphasise various distinctions that subjects may associate importance to or ignore while reasoning.

Further justification may exist in prior work. One may wish to ponder a few rules along with the percentage selection rate of P, Not Q.

| | | |
|---|---|---|
| 19% | ***R17*** | If a card has an 'A' on one side, then it must have a '4' on the other side. |
| 19% | ***R18*** | If a bird has a purple spot underneath each wing, then it must build nests on the ground. |
| 34% | ***R19*** | If a washing label has silk on one side, then it has "dry clean only" on the other side. |
| 34% | ***R20*** | If two objects carry like electrical charges, then they will repel each other. |

Table 1.1: A comparison of four rules that elicit very different response rates of the P, Not Q selections

If one considers the difference between the first two rules and the last two rules all of which were attempted by Cheng et al. (1986), the relation between the antecedent and consequent is clear. Letters and numbers have hardly anything in common, while a washing label and cleaning instructions do. It is possible that rules of this sort are what prompted Relevance Theory (Sperber et al., 1995) and Information Gain (Oaksford and Chater, 1995) to emerge. These theories, however, do not explain the logical nature of the relationship that exists between the two parts of a rule.

Consider the difference exhibited above between the rules R17, R18 and R19, R20. The latter two exhibit a distinction between the antecedent and consequent; in the first, cleaning instructions depend on the material being washed, while they both exist on the same washing label. In the second, the actions to repel or attract are associated with electrical charges and the objects that carry the charges are the same ones that repel or attract. One may then ask, if R18 also speaks about the same bird, in the antecedent it is described to have a spot on its wing and the consequent tells about its nesting habits. However, one must not overlook the fact that the spot on the wing does not relate at all to nesting so it is by no means similar to the charges that repel nor the material that caused the washing instructions to be the way they are.

An issue that can also arise from the analogical stories displayed above is with the surface features of the rules themselves and how they may influence reasoning in the task. A descriptive rule is usually of the form "*If there is an A on one side, then there is a 4 on the other side*" while a deontic rule is of the form "*If the form says entering on one side, then the other side includes cholera on its list of diseases.*" Consequently, the division seemed to be between the abstract nature of having letters like A, B, and Z with having a description of a specific incident that has a "*theme*" as in sending a letter. The abstract therefore is represented by a set of symbols that do not relate to the world except when they are assigned meaning that would cause them to relate to it. For example, Cheng and Holyoak (1985) associated the symbols in the rule "*If one is to take action 'A' then one must satisfy precondition 'P'*" where A is short for Action and P is short for Precondition and they achieved a P, Not Q response percentage of 61%.

This example implies that informing subjects of a relationship between A and 4 may simply be representing a relationship at a high level of abstraction, while A may represent any of "*Action*", "*Aptitude*", "*Assessment*", "*Ace*", etc. Similarly, the number 4 could represent any of "*Plan number*", "*Mark out of 4*", "*value of card*", etc. In this context, their experiment is extremely important, in that it identifies that the relationship imposed is of much higher impact on performance than surface words. However, one may also recall that their response percentage is lower than that which would usually be achieved by contextual questions which reach up to 92%. So although the impact does not seem high, it does not elude a difference that may exist between the two groups of rules at the surface level. Latent Semantic Analysis or LSA (Landauer &Dumais, 1997) is a co-occurrence statistical model that tests how frequently a word appears in the same context of other groups. For example, it would investigate if the number 4 appears next to the words given in the contexts listed above and utilises them to represent it. This analysis, however, is not sufficient because the model takes absolutely no account of the location of the word in the clause because it assumes the two phrases "*one cut*" and "*cut one*" as exactly the same. So a novel tensor based neural network is then utilised to raise the rank of the high dimensional vector space used in LSA by 1 level. Results are expected to be highly informative on one of the possible distinctions between the two groups of rules.

## 1.4 Summary

Chapter 2 reviews only part of the literature that is concerned with the work done here because the amount of work done in this field is truly worthy of respect. This work emphasises a central distinction between thematic tasks that result in a high percentage selections of the desired responses and abstract tasks that result in a low percentage of these responses. Chapter 3 introduces a novel perspective that removes misconceptions about logic and presents the work done to identify the existence of a deontic/descriptive division between the same groups of tasks that result in high percentage selections of desired responses and those that result in a low percentage of these choices.

Chapter 4 introduces the concept of directionality of thought by offering several definitions of different types and follows that with several conveyor experiments. The conveyor setting replaces the anaphora in the task with a unidirectional temporal semantic restriction to identify the influence of directionality on subject behaviour. Chapter 5 utilizes colour as an extra dimension to emphasise various semantic distinctions that subjects may impose onto the task. It also utilises information packaging to alter the hierarchical structure of the sets referenced by the antecedent and consequent.

Chapter 6 identifies through the use of LSA and a novel neural network model that even at that level the order of the words in the rule reflects a basic difference that exists between deontic and descriptive tasks. The latter type exhibits a higher frequency of function words that carry little lexical meaning. Chapter 7 concludes the work done here and suggests many new avenues that can be pursued.

# Chapter 2

# Prior Work in the Selection Task

No tree can grow without a seed, and for the findings made in this work, there were numerous seeds, dispersed over a great forest of work. They come from research made within the selection task itself, research done in linguistics related to the *if .. then..* structure, research done with respect to dynamic reactions, and even philosophical investigations into the essence of language. Last but not least, the most striking finding made in this work represents the effects of the properties of colour on rule semantics. This chapter, however, will take the reader on a tour through the history of relevant work done in this task from the first days of the task through to the current state of affairs. It will start with the shocking results that Wason (1966) found for a seemingly simple task. The journey will then continue to review the investigation methods that range from the classical test to Socratic tutoring. Following this, a need arises to adopt a theory that is capable of explaining behaviour, examples of which include rule familiarity, matching bias, pragmatic schemas and the social contracts. A review follows of the theories of mental representation including Braine's (1978) mental logic and Johnson-Laird and Byrne's (1991, 1993). This is followed by a retrospective review of perspectives, "*egoism*" and Descartes (1954) to conclude with relevance theory and information gain as attempts to formalise results obtained so far by researchers. The review will reflect an urgent need to start looking at the task from a novel perspective that will be presented in the next chapter.

## 2.1 A Logical Shock

The story begins when Peter Wason (1966) presented subjects with a task that seemed logically trivial by giving them the rule "*If a vowel is on one side, then an even number is on the other side*" while presenting them with four cards that have A, B, 4, and 7. They are asked to select the cards they wish to turn over to check the given rule. To the shock of psychologists only between 4% and 20% of subjects made the correct selections of A and 7 (Manktelow and Evans, 1979).

The frantic search for a path towards comprehending why such a simple task caused so much confusion led Wason (1969) to test it in reverse by giving subjects the solution and then the problem. All 20 subjects were able to correctly justify the selections made after only some hesitation while with the original order even the brightest students were not completely convinced of the "*correct*" answers they are given once they complete the task (Wason and Johnson-Laird, 1972). Wason (1969) then tested a variation of the same task that is strictly binary, where subjects are only given one of two possible stimuli as in "*If there is an A on one side, then there is a 4 on the other side*" with cards showing A, B, 4 and 7. Unfortunately, Wason's hopes turned to disappointment because this variation failed to produce better performance. This was followed by a variety of different task formats with rules like *"Every card, which has a red triangle on one side, has a blue circle on the other side"* (Wason, 1969). Results showed a higher selection of P with a lower selection of Q but that reduction was not replaced with NotQ as one would hope. This was followed by another test to identify if subjects were confused by the two sides of the cards, in particular whether or not they interpreted "*the other side of the card*" to imply the side meaning the side that faced "*downwards*" only. All the information was presented on the same side of the card with a part of the information hidden behind a mask, but this too was regarded as a big disappointment (Wason and Johnson-Laird, 1970). Negation came to mind and instructions that asked subjects to pick the cards that 'could break the rule' were used, hoping to raise the percentage selection of NotQ. Another disappointment resulted because it only induced subjects to select NotP and NotQ conditions rather than the correct P and NotQ cards.

> *"Most of these experimental modifications could be regarded as 'failures' because they did not improve the subjects' performance (Wason and Johnson-Laird, 1972)."*

Researchers regarded the results at that time almost unanimously as a source of embarrassment because subjects are not thinking in a logical way. Stenning and Lambalgen (2002) however, now inform us that we are not supposed to jump to such a conclusion without looking at what we consider a 'correct' answer to the task. Wason and Johnson-Laird (1972) assure the reader that it is a simple task.

> *"It may be hard for the reader, guided by our concrete example, to appreciate the extreme difficulty of the task. (If he wants a proof of it, let him try it out on his own friends." Time after time our subjects fall into error. Even some professional logicians have been known to err in an embarrassing fashion, and only the rare individual takes up by surprise and gets it right. It is impossible to predict who he will be. This was all very puzzling and, if anything, made it harder to understand the origins of scientific enterprise. But we wondered whether our subjects' difficulties were due to the very simplicity of the problem. It is really simple."*

Stenning and Lambagen (2002) on the other hand, now wonder if the task is as simple as we were led to believe and wonder if one is to automatically 'expect' subjects to think using classical logic, without any regard given to the semantics of the premises in the rules. In other words, are we really asking the right question and is the normative answer we are checking answers against, the correct one?

Wason (1987) shows that he has a strong belief that subjects are guided by old prejudices that govern the way they respond to this task where a fact is not sufficient to alter one's beliefs. It is ironic that Wason is displaying a type of confusion in explaining the task that is no different from that which is displayed by subjects. What subjects fail to do is to check the possibility of having a false consequent of the If P then Q rule which is NotQ. To identify the similarity between this and what researchers are doing we will examine Popper's (1963) philosophy of scientific discovery.

Popper (1963) starts his theory by examining the sources of theories through investigating several possibilities. Assumptions that truth exists and is clear to all those who seek it is clearly untrue as we have examples where Einstein corrected Newton's theory even though it produced accurate results for all studies done on earth within our limited 3-D world. Anomalies existed in outer space and the ignorance of scientists at the time does not imply that they are not scientifically relevant. Therefore, existing theories may be affected by prejudices that many philosophers seek to purify themselves from when seeking true knowledge. Based on this he defines the conspiracy theory of ignorance as the possibility of one to err as a result of the influence of prejudices imposed through tradition, social norms and the like. He also explains that erroneous theories may exist for thousands of years as the belief that related illness to evil forces in a field that is currently regarded as medicine.

Therefore, erroneous theories await refutation and one possible route towards this is to question the sources of these theories by questioning the assertions they are based on. However, he shows that if this is done, then it is liable to lead to an infinite regress and cast doubt on all existing theories as it does not question whether the theory holds any "*truth*". All it actually questions is the adequacy of the sources of ideas and these sources are obtained through either senses that err or erroneous misinterpretations depending on the philosophy adopted to explain human error. With the possibility that a "*true*" theory can emerge from faulty assumptions, this questioning of sources becomes futile and leads only to infinite regress. He comes to the conclusion that:

> *"For if we are doubtful about an assertion, then the normal procedure is to test it, rather than to ask for its sources; and if we find independent corroboration, then we shall often accept the assertion without bothering at all about sources." (Popper, 1963, p.23)*

This theory clearly allows premises to be considered as true provided one seeks to refute the given assumptions. If it applied to the selection task rule, we find a rule such as "*If A is on one side, then 4 is on the other side*" can be considered as being held provided it is not refuted. The rule is first checked through checking the A card, following with a sincere attempt to find a false instance by checking the 7 card. Several researchers including Stenning and van Lambalgen (2002) and Oaksford and Chater (1995) have noticed this analogy between this theory and the task. The basic assumption made by Popper (1963) can be written in rule format as "*If we are to accept a theory, then all findings must support it.*" So to check this rule, scientists need to engage in finding new theories and then seeking to falsify these theories allowing a theory to continue to exist only if it survives these tests. No one should object that this is indeed the claim made by Popper (1963). This requirement is perfectly aligned with the "*expectations*" of most psychologists working on the task to check the cases of P = "*accept a new theory*" and NotQ = "*Check if findings do NOT support it*".

So what are scientists actually doing with this task? Oddly enough it seems that scientists are making the same error their own subjects are making because they are not attempting to falsify their theories that attempt to explain why subjects do not select the NotQ falsifying card. Wason (1987) indicates that subject responses reflect that they find a great difficulty in getting rid of old prejudices like 'the earth is flat' in the face of new evidence. He explains that Kuhn showed that old paradigms continue to reign even if a few false instances arise. Therefore, it is clear that Wason "*expects*" subjects to check the false consequent even though he himself has fallen prey to old prejudices when he claims that there are exceptions to theories. If there are exceptions then to what gain is checking the NotQ card? Others followed closely in his tracks assuming that verifying their theories is sufficient for their theories to hold without attempting to falsify them, thereby repeating the exact "*error*" they are investigating.

Consequently, someone must break the cycle by introducing a new method of analysis or approach to the problem. The first spark originated from the same person who found the problem, Peter Wason.

## 2.2 Socratic Tutoring

The Socratic method of inquiry was originally introduced to help discover truth and is composed of a series of questions that a skilful teacher utilizes to "*guide*" a student to find the path to "*correct*" reasoning. This form is based on allowing a person to indicate a belief X, and following a series of question to check if the person also believes Y. If yes, then that person is led through discussion to agree that Y implies not-X and this highlights a contradiction that in effect forces the person to decide if he prefers X or Y. (Ross, 1996)

Wason and Johnson-Laird (Wason and Johnson-Laird, 1970) were the first to utilise the Socratic method to engage subjects into a dialogue and lead them into a state of "*conceptual conflict*" as described by Wason and Johnson-Laird (1972). Their aim was to identify the "*path*" of thought and to raise the possibility of conceptual conflict as a subject arrives at the state of contradiction to be able to "*see*" the benefit of selecting the NotQ card. They gave subjects the cards and question and through discussion attempted to identify the types of conflict that may arise during the reasoning process.

As a result, they isolated two types of conflict; hypothetical and concrete contradiction. They defined a hypothetical contradiction as when a subject recognises that a possible value of P on the other side of the NotQ card would falsify the rule, yet these subjects still refrain from selecting that card. In this case, subjects seem "*aware*" that a possible falsifying instance may exist, and still have no desire to check that instance for unknown reasons. However the authors deny that it is possible that subjects are actually 'aware' and select the wrong answers.

A concrete contradiction, on the other hand, occurs when the cards are actually flipped and when a card with P is flipped to reveal a Q, and this followed with a card of NotQ flipped to reveal a P. The first fits the rule while the second contradicts the rule leaving subjects with a concrete contradiction. Wason and Johnson-Laird (1972) indicate that the conceptual clash is obvious and explicit.

They wonder why subjects seem persistent in their choices in spite of the fact that they are sometimes aware that Q for example is useless, and yet they select to turn it. They arrive at the conclusion that in spite of a clear confusion that subjects seem to face, the real reason behind this behaviour is that they are compelled to "*resolve two seemingly irreconcilable thought processes*". This raises the issue that the two "*correct*" choices according to the classical logic metric may truly be flawed. In fact, Wason (1987) clearly indicates that subjects "*will frequently deny the facts, or contradict themselves, rather than shift their frame of reference*." Here we find incredibly strong support for the claims made in this thesis to the effect of "*directionality of thought*". There seems to be a strong bias that dictates the lines along which the deduction process is able to proceed, evidently from the initial days of the emergence of this task.

Unfortunately, although this method of dialogue was extremely fruitful in the limited domain to which it was applied, it was not replicated on a wider scale as a probe into subject thought processes. If we consider the group of tasks classified as successful "*thematic*" tasks, for example, we find no Socratic dialogue studies of student reactions to encountering conflict of card contents to the rule. Perhaps a possible reason is that with this approach, subjects were actually performing an evaluation task because they would actually turn the cards over and evaluate the other side, while in the original task all they have to do is select the cards they wish to turn over. It is well known that this difference in the task is usually sufficient to make the two experimental formats different enough to cast doubt on whether results are comparable, and consequently most researchers avoided it. Unfortunately, the path followed displays results of question formats while keeping subjects' reasoning methods ambiguous.

Therefore, the search was on, for the theory that would offer its creator fame and fortune through its ability to explain subject behaviour in the task. In short, it became a guessing game that is not "*enlightened*" through tools such as Socratic Tutoring.

## 2.3 In Search of a Theory

As a result of the extreme shock arising from subject behaviour, a frantic search for a theory to explain behaviour in the task was immediately underway. The initial motivation was to find what materials would result in better performance and since the authors provided no insight as to the possible source of these theories, an attempt will be made to explain them while trying to link similar efforts to each other.

### 2.3.1 Familiarity of Rule

The initial tests that showed the effect of "thematic" materials included those by Wason and Shapiro (1971) indicating a possible facilitating effect of certain types of materials. This led to the possibility that the "*familiarity*" of the rule is a contributing factor to good performance. In simple terms, a rule is familiar if subjects have prior knowledge of the rule and know some possible falsifying instances. One example of this is when Johnson-Laird, Legrenzi and Legrenzi (1972) tested a postal rule on British subjects where the percentage of correct responses was 81%. On the other hand, Griggs and Cox (1982) failed to observe facilitation by the same rule on American subjects with the rule "*If a letter is sealed, then it has a 15-cent stamp on it*". Golding (1981) later found that the same postal rule produced better results when attempted by older British subjects who were familiar with a regulation that is no longer imposed by the British postal office. It seems that subject familiarity with the rule alerts subjects in some way to the possibility that a false instance of the rule may occur. However, Manktelow and Evans (1979) indicate that rule 'familiarity' interacts with some other factors and is not a straightforward issue to isolate for the benefit of predicting better performance.

Therefore, although 'familiarity' seems intuitive, it places an extremely broad range of rules into the group and does not highlight what aspect of familiar rules actually influences the task. One example of this is, whether the familiarity must simply involve the possibility of cheating on the rule, or whether it must it be a personal experience familiarity or other possibilities. Therefore, all this theory seems to do is to claim that when background information intervenes in a "thematic" task, in some cases, subjects will do better. It offers no information as to the specific semantic requirements of the background information that is necessary to interact with the task in order to result in facilitation. Consequently this approach did not prove to be extremely beneficial in predicting performance because it does not answer the questions that arise as to why subjects do not select the "*expected*" answers and why they seem confused with the abstract task. It does not even give enough information to help predict what background information is necessary to influence subject responses in the task. In retrospect, the description is insufficient to explain subject behaviour when they select P and NotQ or any of the other selections subjects make.

### 2.3.2 Matching Bias and Negations

One of the first theories that set out to explain why most subjects select P and Q cards is the matching bias theory as introduced by Evans (1973). It explains behaviour by assuming that subjects choose the cards, which "*match*" the antecedent and consequent clauses of the rule, while they ignore the negations that may be present, especially that of the consequent. So, if a subject sees a rule of the form If P then Q, that subject is more likely to select P and Q rather than select P and NotQ. It is self evident that this theory describes the phenomenon that a high number of subjects select P and Q. But with matching bias, if a subject has a rule "*If the letter is a D, then the number is not a 4*", this subject would select D and 4 which are the correct choices for this rule. In other words, they would match the named values, not their truth values (Evans, 1993). However, this theory only explains the behaviour of a group of subjects that actually make the selections of P,Q and ignores all others that form the majority and are dispersed amongst other possible choices.

> *"Theories of reasoning based on formal rules propose that the ability to make suppositions is central to deductive reasoning." (Handley and Evans, 2000)*

Evans et al. (1998) then modified the original theory by explaining how subjects process the two types of reasoning: Modus Ponens and Modus Tollens. Modus Ponens is when a conditional premise is given with a fact as follows: "*If P then Q, P*". A subject utilises *fact P* and the rule *If P then Q* to deduce *Q*. Modus Tollens is when a conditional premise is given with the negative of the consequent as in "*If P then Q, NotQ*". In this case, they inform us that a subject must "*suppose*" that P is true and from that deduce *Q*, following which the subject would arrive at a contradiction with both Q and NotQ as true. Once this contradiction arises the subject would use a reductio ad absurdum rule to deduce NotP and they indicate that this explanation is the one provided by mental logicians. We are, however, not informed when or why subjects utilize each of these two types nor are we informed why we can assume that all subjects follow the same path.

Additionally, we are informed that selections inform on the perceived "*relevance*" of the cards, which is determined through an "*if-heuristic*" and a "*not-heuristic*". The if-heuristic leads a subject to imagine a hypothetical world where the antecedent of the rule is true, and this, we are told, is the reason subjects select P more frequently than Q. The not-heuristic, on the other hand, implies that "*a negated statement is a comment on the proposition it denies*" and because it is a comment, subjects select the named value alone without the negation. Evans et al. (1998) offer additional support for the claim that subjects face a serious problem in recognising that Not-Not-P is the same as P, such that the double negation effect is suppressed if the conclusion is ahead of the premises or if the minor premise is ahead of the conditional.

In retrospect, this theory is extremely interesting with respect to the information it adds through testing negative instances of the premises. However, it does not offer any explanation as to why subjects are distributed amongst various possible choices selecting either P alone, or selecting P, Q and Not Q together while others select all the cards. Therefore, the theory offers no explanation of all behaviour that is observed to result in the task.

It does, however, raise the issue of the inability of subjects to equate Not-Not-P with P. This places negatives into the limelight, making us wonder what NotP represents to our conceptual system. One may consider, for instance, that while a double negative becomes a positive, there is no such thing as a double positive. Additionally, it is obvious that the negation of a premise in the abstract questions does not seem to affect the value of the premise as, in matching bias, subjects would select that premise without its negation. In a conditional of the form "*If A then Not 7*" the "*7*" is explicitly displayed in spite of it being next to a "*Not*". Perhaps the neglect of this "*not*" is what led Evans (1993) to describe negation as a "*comment*". However, this description cannot possibly exist in the same conceptual system that has problems with a double negative and gets confused. Wason (1987) offered another explanation as to why subjects fail to select the NotQ in the usual order of the task as a difficulty that subjects face when "*shifting their frame of reference*". Although nothing was given to indicate if this is supposed to include the cases where the rule is If P then NotQ, it does seem appropriate for this case. In the given rule a negative of the consequent is given, making it unnecessary for the subject to shift their frame of reference to obtain the correct Q selection. A possible reason for this could be that the explicit NotQ according to Sperber et al. (1995) is logically more difficult to represent than an implicit value other than Q that reflects

NotQ. An example is like comparing Not Black to White, since while both are contraries of Black the first is more difficult to arrive at, so arriving at Black from the first would then be cognitively easier than arriving at Black from the second.

Another way of looking at Wason's (1987) frame of reference, therefore is to look at two mirrors facing each other. We have a foreground rule of the form "*If A is one side then 4 is on the other side*" where A belongs to the set of letters while 4 belongs to the set of numbers. If we reflect on this rule through a frame of reference that reflects on its contents then it is similar to looking at it through a mirror. The rule therefore is the relationship between these two particular values from these sets. We also have a background rule that states that letters are on one side of the card while numbers are on the other side of the card. Additionally, there is an implicit relationship between the two letters A and B such that whenever we see B we are to assume that it is a "*not A*" and similarly when we see a 7 we are to assume it is a "*not 4*". This implies a second reflection such that each value in the rule has a "not" of its truth. The implicit rule is that "*If not 7 then not A*" and this has to be true to ensure that the first rule is valid. In a mirror we see our right hand as left and our left hand as right. In a comparable manner this implicit rule reflects the first while exchanging the first and second premises. Consequently, it should not be surprising that if subjects regard the rule through Wason's frame of reference (1987) shifting from one to the other is comparable to seeing what is in one mirror through another mirror because each would exchange the order of the premises leading subjects to infinite regress. In the following section, social contracts will avoid this problem by explicitly giving subjects the ability to follow a single line of thought to both selections. They will also highlight another issue regarding perspectives.

### 2.3.3 Pragmatic Schemas Lead to Social Contracts

Pragmatic Reasoning Schemas is a theory offered by Cheng and Holyoak (1985) based on the successes of "thematic" rules. The theory describes a type of knowledge structure that facilitates reasoning in this task. This type of schema consists of a set of generalized context sensitive rules which are defined in terms of pragmatic goals that guide the process of inference. Therefore, each subject is given a goal through the context of the question and the means to arrive at that goal through card selection. These schemas include the "*permission schema*", the "*obligations schema*" and "*causations*". Context sensitive rules are then defined by the conceptual system in terms of the goals given in the question. These rules include taking desirable actions, making predictions about future events as well as defining relationships to these goals as in cause and effect or precondition and allowed action. Additionally, the framework assumes the role of familiarity or prior experience in facilitation of the "*correct*" answers to the task.

Cheng and Holyoak (1985) explain that the abstract task does not evoke any schema because it is not related to any life experiences, while most of the successful "thematic" problems fit a "*permission schema*" format. Inference patterns include the concepts of possibility, necessity, an action to be taken and a precondition to be satisfied. We are also informed that this is parallel to the deontic concepts of permission and obligation as expressed by *can* and *must*. The core of the permission schema is then summarised into four production rules as follows:

> "*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, 1985, p. 397)*

They explain that if a subject is given a permission scenario the whole set of rules is invoked. Rule 1 has the same effect as Modus Ponens, while Rule 2 blocks the fallacy of Denying the Antecedent. This fallacy occurs when a subject is given the rule *If P then Q* and then given NotP. Subjects erroneously jump to the conclusion that *NotQ*. Rule 3 blocks the fallacy of Affirming the Consequent, which occurs when subjects who are given the rule *If P then Q* and then given *Q*, erroneously deduce *P*. Rule 4 represents the contrapositive of the given rule, stating that if the precondition is not satisfied then the action must not be taken.

They also indicate that permission schemas are context sensitive and directly related to deontic concepts such as *must* and *may* that cannot be expressed in standard propositional logic. However, they point out that the rules are heuristic rather than logically valid inferences. An example is that Rule 3 does not always follow from Rule 1. The problem arises when an action requires more than one precondition. For example, "*if one is to rent a car, then he must have a driving licence*" could represent Rule 1. However, there may also exist another requirement such as that person must not be under 21 years old. Therefore, Rule 3 that results from the above would not hold as it would be "*if one has a driving licence, then he may rent a car.*"

They also indicate that a causal schema may invite subjects to support the converse of the given conditional. So a causal conditional of the form If *P then Q* where P = "*someone let go of a ball*" causes Q = "*the ball falls to the ground*", can lead to the assumption that If Q = "*the ball falls to the ground*" then P = "*someone let go of the ball*" because one assumes that the Q consequent results when the P cause is true. This conclusion is not possible through classical logic inferences.

This claim is supported through several experiments, one of which tests the rule: "*If one is to take action A, then one must first satisfy precondition P".* The cards then show the possibilities: *has taken action A*, *has not taken action A*, *has fulfilled precondition P*, *has not fulfilled precondition P*. Subjects are cued into the role that they are an authority checking whether or not people are obeying regulations. Results came in confirming the predictions made by the theory when a percentage of 61% of the subjects solved the schema version correctly.

In retrospect, there are clear indications that this theory has strong links to deontic logic because permissions and obligations imply the existence of an ideal state, while causation can be compared to obligations. In permissions, an action must not ideally be taken unless proper requirements for it have been satisfied. In obligations, if a particular situation exists then one is obliged to perform a duty.

Additionally, causation is similar to obligations except that it simply describes a natural follow up of one state to another showing an expected sequence of events. The heuristic nature of the theory shows that it is not strictly "*logical*" in the classical logic sense and is instead closer to deontic logic which the authors explicitly acknowledge.

An especially interesting aspect of the experiment described above is that it was one of the first to show a sequence effect of one "*type*" of task on another "*type*", namely a permission rule and an abstract card question. They noticed that the card problem was solved more correctly 39% when it followed the permission problem, than when it preceded it only with 19%. In converse, the permission problem was solved more correctly when it came first 61% than when it followed the abstract problem 48%. In prior tests done by other researchers this effect was not present, which implies that the "*permission*" type question they used has special features that allow it to facilitate performance in the abstract task and by converse be adversely affected by it. This issue is further analyzed in this work in Chapter 4.

Another theory that exists in close competition with the above was introduced by Cosmides (1989) namely; Social Contracts. She claims that natural selection theory forms the basis for this theory because it is regarded as informative of the kind of cognitive psychological mechanisms that are likely to be made universal for all conceptual systems as with the behaviour exhibited here. This, we are told, restricts the range of possible culprits to the results that are observed. We are also informed that the theory of natural selection is responsible for pinpointing adaptive problems that the human mind must be able to solve, as well as suggesting possible new mental design mechanisms that may be necessary. On the other hand, evolutionary biology, we are told, gives a high level definition of how processing can actually be achieved and it dictates adaptive goals for the brain that must be realised. The theory is also based on the assumption that the innate cognitive architecture is composed of a large array of highly specialised adaptive mechanisms. Evidence for such mechanisms is then reported to be that reasoning performance is content dependent and that the performance is altered by specific content in the predicted adaptive direction.

The author links this to the content dependency of the selection task results to date. She indicates that although pragmatic schemas as defined by Cheng and Holyoak (1985) are content dependent, their rules are created by content independent cognitive processes. She indicates that their permission problem that is shown above is not compatible with the abstract problem it is compared to. Evidently the permission question contains the paraphrase: "*In other words, in order to be permitted to do "A", one must first have fulfilled prerequisite "P""* (Cheng and Holyoak, 1985, p. 403). No analogous text exists in the abstract question which causes Cosmides (1989) to cast doubt on the actual cause for the difference in performance.

She defines a social contract as one that requires subjects to judge between <u>perceived benefits</u> and <u>perceived costs</u>. This presumption is then tested through the comparison of two contexts: the first "*a social contract*", while the other is a descriptive context. The social contract associates the task to be done with a clear benefit that is to be weighed against a clear cost, while the descriptive task does not define terms as costs or benefits but does link them by a familiar relation.[3] She reported a clear difference in performance between the two versions with 75% correct for the social contract and with 21% correct for the descriptive problems.

---

[3] The word 'descriptive context' here should not be confused with the word 'descriptive' used in the deontic context because here the definition is given strictly according to the one used by Cosmides while there the definition is based on deontic logic.

Cosmides (1989) makes an extremely interesting comparison between the standard tasks and what she described as a private exchange with the form "*If you do X for me, then I'll do Y for you*". She tested this version in experiment 3 with the standard order of the rule as in "*If a man eats cassava root, then he must have a tattoo on his face*" with correct answers *P* and *NotQ*, and in experiment 4 with the reverse order of the rule as in "*If a man has tattoo on his face, then he eats cassava root*" with correct answers *NotP* and *Q*.

However, before one can go into the analysis of the results of this experiment, an issue of a central distinction between cheater detection in social contracts and logic inference has been raised by Stenning and van Lambalgen, (2004). They indicate that Cosmides regards the two rules shown above as equivalent and both would lead to the choices of "*man eats cassava root*" and "*man does not have a tattoo on his face*". In the case of the first rule, these would represent the choices of P, Not Q while for the second rule they would represent the choices of Not P and Q.

From a logic perspective this is not the case because the premises are regarded such that it is necessary for the first to be true so that the second is true, which imposes an order on the rule. Consequently, from a logic point of view, we find that the answers to the first rule would be "*man eats cassava root*" and "*man does not have a tattoo on his face*" which are P, Not Q while for the second rule the answers would be "*man has a tattoo on his face*" and "*man does not eat cassava root*" which are P, and Not Q for the second rule.

Now to analyse the influence of the private exchange and how it interacts with rule order, we find that in the standard order there was no clear difference between the standard task and the private exchange. However, we find a curious emerging behaviour when Cosmides tests the reverse order in that subjects make what she regards as a "*mistake*" by selecting P and NotQ when she expects the correct answers to be NotP and Q. It should be clear from the argument above that the first choices are not at all a mistake but instead are in line with what logic would lead them to select as opposed to what social contracts would dictate.

Consequently, the effects of having a "*private exchange*" on these particular selections would not only show that subjects are sensitive to private exchanges, but it would also relate this sensitivity to how they reason logically. In fact, we find that the unfamiliar descriptive task showed an increase in selections of the two cards of P and NotQ from

12% to 25% by the private exchange version. The really large difference, though, was exemplified in the abstract card version where P and NotQ selections rose from 12% in the reverse order standard to 33% correct in the abstract reversed private exchange. With N = 24 for both experiments and assuming that the private exchange will have no effect on the reverse order of the task, so the standard value of the reverse order is taken as expected, we get a chi-square value of 9.524 and $\underline{p} < 0.002$. The private exchange exhibited here includes the use of the words "*I*" and "*you*". This seems to present itself as a strong indicator that the direction of the rule interacts with these words such that they affect behaviour.

In an attempt to counter criticisms, Gigerenzer and Hug (1992) provide further support through experiments to fortify Cosmides' stand by clarifying the main differences between Pragmatic Reasoning Schemas and Social Contracts. They explain that while all social contracts can be regarded as pragmatic schemas, the converse is not true as some pragmatic schema formats can fail to obtain good subject performance. They emphasise the importance of the fact that a subject has to be cued into "*the perspective of a party that could be cheated*" and indicate that this can imply that subjects could be cued into two opposite perspectives. A sample rule of this could be: "*If an employee works on the weekend, then that person gets a day off during the week*". Half the subjects are given an "*employee*" version of the problem and informed about an employee who never worked on the weekend before but is considering doing that from time to time, since having a day off during the week is a benefit that outweighs the cost of working on Saturdays. There are rumours that the rule has been violated before. Subjects must check whether there is a case where an employee worked during the weekend and did not get a day off during the week, so their correct selections would be P and NotQ. The other half of the subjects, are given an "*employer*" version of the problem and given the same rationale above. These are cued into the perspective of the employer who must find out if there are employees that do not work on the weekend, and still take a day off during the week which are NotP and Q.

Yet again we find evidence to support the claim that the alignment of the expected answers namely P, NotQ gives a percentage of 71% which is higher than the reverse perspective which expects NotP, Q as answers and has attained a percentage of 61%. Differences are also evident in the other two rules that were tested with 71% vs. 64% and 81% vs. 59%. The frame of reference a subject assumes and reasons for it seems to have a strong effect on all cases of the task, especially the deontic versions of it. What is still unclear, though, is if it has any bearing on the abstract classical task. However, before one could delve into any of this, a very brief review of theories of mental representation that have been applied to this problem claims our attention.

## 2.4 Braine's Mental Logic

At this point in time, it should be interesting to regard the analysis of the results in the task from a logician's standpoint, but this approach just attempts to passively explain results in terms of a "*mental logic*" such that subject behaviour is nothing more than an "*error*". Consequently, mental logicians do not make the same arguments made by general logicians and cannot be regarded as representing the latter.

A prominent example of this approach is Braine's abstract rule theory (Brain, 1978; Braine & O'Brian, 1991; Brain and Rumain, 1983; Brain, Reiser & Rumain, 1984; O'Brian, Braine & Yang, 1994; Rumain, Connell & Braine, 1983). The theory starts off with an assumption that there are two types of reasoning: formal reasoning and practical reasoning. The latter is the one used by people in their daily lives because they are capable of extracting information from experience to aid them in their deductions, while in formal reasoning one has to be restricted by what is given. Additionally, the theory identifies three types of errors that can divert the deduction process from its path. Comprehension errors occur when the interpretation of premises or assumptions is faulty or inaccurate. Heuristic inadequacy errors occur when the conclusion to a reasoning problem cannot be reached because the strategies used to select the rules are inadequate. Processing errors result from lapses of attention or failure to take account of all relevant information in working memory.

The practical logical process is initiated when the premises are first encoded by some mechanism and the resulting representation is related to abstract reasoning rules. In a way, this is like uncovering the logical form of the sentence except that we are informed that *If P Then Q* is *noncommittal* about the basis for the inference rule it states. If a speaker uses that phrase, then the speaker is giving the listeners a reason for concluding q given p but it provides no reason why that should be so. A listener may infer that p is the cause of q, that p entails q, or that q follows from p together with shared assumptions or that there is some correlation of unknown basis between the events. In other words this is subject to the errors described above as well as showing a degree of ambiguity between the logical form of the sentence and the sentence it was extracted from.

These rules are then used as basic rules from which to draw valid conclusions as per the inference rule schemata. Rules would be selected and applied whenever needed to produce a chain of inferences that would form a mental proof of derivation to a conclusion. If this process does not deliver a straightforward conclusion, then a set of non-logical rules are utilized to determine the response resulting in biased responses.

The theory identifies some problems that logic theories suffer from in general, describing one as the problem of directionality difference (Braine, 1978). In a rule of the form *If P then Q*, then we can logically assume that *If NotQ then NotP*. However, we are informed that due to effects of rule directionality on thought, the two are not cognitively equivalent. This claim is supported through three sentences where each is given in the formats *if ..then.., ..only if.., and if not .. then not ..*

> *"3a If one pulls out the knob, the television goes on.*
> *4a If the triangles ABC and PQR are congruent, then AB = PQ.*
> *5a If the television goes on, then the knob has been pulled out.*
>
> *3b One pulls out the knob only if the television goes on.*
> *4b The triangles ABC and PQR are congruent only if AB = PQ*
> *5b The television goes on only if the knob has been pulled out.*
>
> *3c If the television does not go on, then one does not pull out the knob.*
> *4c If AB≠PQ then the triangles ABC and PQR are incongruent.*
> *5c If the knob has not been pulled out, then the television does not go on."*

Braine (1978) informs us that the first sentence **3a** is a causal relation, the second **4a** an entailment and the third **5a** a relation in which the antecedent is evidence for the consequent. Note that **4b** and **4c** both make sense, but they differ in meaning when aligned with **4a**. In **4a** the consequence is a deduction from the antecedent while in the other two it is a condition of its truth. A comparison between 4a and 4c would show a direction reversal between the two, yet their truth tables are identical. He goes on to give another example of the importance of directionality in the conditional that exists when *if and only if* is equated with $p \equiv q$ which logically means the same as $q \equiv p$. However, the sentence "*They went to the party if and only if they were invited to it*" does not mean the same as "*They were invited to the party if and only if they went to it*", even though they may have the same truth conditions under some interpretations depending on the temporal reference.

Braine (1978) also reported two experiments to test for the effect of directionality on the common errors in reasoning where subjects were presented with four types of problems several times. The four types are: Modus Ponens, Modus Tollens, Denying the Antecedent, and Affirmation of the consequent.

| | |
|---|---|
| *Modus Ponens*: | If P then Q<br>P<br>Therefore, Q |
| *Modus Tollens*: | If P then Q<br>Not Q<br>Therefore, NotP |
| *Denying the Antecedent*: | If P then Q<br>Not P<br>Therefore, Not Q |
| *Affirmation of the Consequent*: | If P then Q<br>Q<br>Therefore, P |

In one half of the presentations the rule had the form *If P then Q,* and in the other half it had the form *Q only if P*. When the rule is in the form *If P then Q* then Modus Ponens is the easiest to apply and the errors of Denying the Antecedent and Affirmation of the Consequent disappear. However, when the fact is something other than P, things get more complicated. If the rule presented is of the form *Q only if P* and NotP is given then the rule format makes it easier to deduce Not Q.

Results confirmed these predictions by showing that when the first form is used *If P then Q*, Tollens had significantly more errors than Ponens. By contrast with the second form *Q only if P* Tollens had significantly less errors than Ponens. Braine (1978) regards this result as a clear indication of the directionality of the inference rule.

Therefore, although this theory may explain low accuracy results in the Wason task (1966) by accepting them as errors, it does highlight the importance of directionality to an inference rule. Understanding this importance is taken forward an extra step in the work presented here.

## 2.5 Insight from Mental Models

Although this theory was not designed to resolve the problems of this task, it is interesting to consider its input and that of its creators as to the "*effective*" manipulations made to this task. The Mental Models theory (Johnson-Laird & Byrne, 1991, 1993) explains behaviour in the task through an assumption of a representational system that generates mental '*models*' to represent the problem at hand, based upon the "*truth*" of premises as well as general knowledge. The claim is that it is based on the Principle of Truth.

> *"A fundamental assumption of the theory is the principle of truth: Individuals minimize the load on working memory by tending to construct mental models that represent explicitly only what is true, and not what is false" (Johnson-Laird, 1999)*

Deductive reasoning is assumed to involve three processes: the comprehension of premises to form a model, or set of models, the combining and description of the models to produce a conclusion, and the validation of this conclusion by eliminating alternative models that do not result in the expected conclusion. During the comprehension stage, the models are formed in a way similar to this example: First a subject is given this rule, and that is followed by a fact, from which the student is asked to make a conclusion.

| | |
|---|---|
| *If it is raining, then Alicia gets wet,* | *If P then Q,* |
| *Alicia is wet.* | *Q,* |

The principal of truth is applied at two main levels: the first level is where subjects represent only true possibilities and the second to represent for each of the possibilities the literal propositions that are true whether affirmative or negative.

Then according to the theory, three models are formed:

| | | |
|---|---|---|
| *1.* | *It is raining* | *Alicia gets wet* |
| *2.* | *It is not raining* | *Alicia gets wet* |
| *3.* | *It is not raining* | *Alicia does not get wet* |

However, Johnson-Laird and Byrne (1991) suggest that when people understand this premise, they start by building the following initial models:

*[It is raining]* *Alicia gets wet*
…

The square brackets mean that the premise has been represented "*exhaustively*" in the model so in a model that has this premise there must also be the second premise which is "*Alicia gets wet*". The three dots indicate the existence of more models that can be obtained from the same premise. When the sentence "*Alicia is wet*" is given, it is just matched onto the shown model to arrive at the conclusion that "*It is raining*" without the need to flesh out any more models.

Johnson-Laird and Byrne (1991) describe subject behaviour through the Mental Model theory as follows:

1. Subjects only consider the cards that are explicitly represented in their models of the rule.

2. They then select a card for which the hidden value could have a bearing on the truth or falsity of the rule.

A rule such as *If there is an A then there is a 2* would yield the following models:

*[A]* *2*
…

Subjects will consider both cards but will select only the "*A*" card because it alone has a hidden value that could affect the truth or falsity of the rule. If on the other hand subjects interpret the rule as a bi-conditional the following models would result:

*[A]* *[2]*
…

This will make them select both the cards "*A*" and "*2*". If on the other hand, an *independent* bias exists indicating that the rule may be broken, then the following models would result:

*[A]* *2*
¬*2*

This makes it extremely difficult if not impossible to arrive at the following model given the abstract rule unless a clear indication is there to lead subjects to deduce not 2:

*[A]* ┐*2*

They claim that this prediction makes sense of the five experimental manipulations that have been found to yield the correct selections, which are as follows:

### 2.5.1 Changing the form of the rule as in "*only if*" formulations should enhance performance.

Notice that Johnson-Laird and Byrne (1991) dismiss the idea of "*directionality*" and justify the results obtained by Braine (1978) by showing the models that result from the two rule forms:

| *If P then Q* | | *P only if Q* | |
|---|---|---|---|
| P | Q | [P] | Q |
| [P] | [Q] | ┐P | [┐Q] |
| | | … | |

In the case of *P only if Q*, the models allow both Modus Ponens and Modus Tollens to be made without any further fleshing out, which is not the case with regular form. However, this prediction is not supported by the findings of Cheng and Holyoak (1985) who found that reasoning with If P then Q is easier than reasoning with P only if Q. Note that the emphasis on the effect of directionality persists because the difference in performance continues to exist.

### 2.5.2 Changing the content of the rule through utilizing contents that trigger memory for times when the rule was violated.

Griggs and Cox (1982) presented subjects with the following "thematic" question:

> *"On this task imagine that you are a police officer on duty. It is your job to ensure that people conform to certain rules. The cards in front of*

> *you have information about four people sitting at a table. On one side of the card is a person's age and on the other side of the card is what the person is drinking. Here is a rule: If a person is drinking beer, then the person must be over 19 years of age. Select the card or cards that you definitely need to turn over to determine whether or not the people are violating the rule."(p. 415)*

Four cards were presented to subjects containing the following: "*Drinking a beer*", "*Drinking coke*", "*16 years of age*" and "*22 years of age*". This is a clear example of a deontic rule where a falsifying instance does not cause one to doubt the existence of the rule. Additionally, consider this rule: "*If a person is over 19 years of age, then that person must be drinking beer*." The first thing that comes to mind is a violation of that rule, because the implications it has simply by changing the order of the premises are enough to show that even if a person is old enough, this does not mean that the person is "*drinking beer*", either not drinking at all, or not drinking at this moment. Notice the effect of tenses in this order and the strong effect of order. In a sense, it shows how strong this actual rule is with respect to the directional bias it imposes on subjects and its sensitivity to verb tenses.

### 2.5.3 Changing the context of the rule as in the pragmatic reasoning schemas and social contract theories leading to the explicit representation of negative instances.

If the context implies a "*role*" as in pragmatic reasoning schemas or presents a social contract when one has to weigh benefits against costs, negative instances or NotQ cases are explicitly highlighted raising the possibility that they may be true. An example of such a rule along with the models that represent it follows:

> The rule is: *If a man has a tattoo on his face then he eats cassava root*.

Subjects are given the following information:

> "*You are an anthropologist studying the Kaluame, a Polynesian people who live in small warring bands on Maku Island in the Pacific. You are interested in how Kaluame "big men" – chieftans – wield power.*
>
> *Big Kiku is a Kaluame big man who is known for his ruthlessness. As a sign of loyalty, he makes his own subjects put a tattoo on their face. Members of other Kaluame bands never have facial tattoos. Big Kiku has made so many enemies in other Kaluame bands, that being*

*caught in another village with a facial tattoo is, quite literally, the kiss of death.*

*Four men from different bands stumble into Big Kiku's village, starving and desperate. They have been kicked out of their respective villages for various misdeeds, and have come to Big Kiku because they need food badly. Big Kiku offers each of them the following deal:*

*"If you get a tattoo on your face, then I'll give you cassava root.*"

*Cassava root is a very sustaining food which Big Kiku's people cultivate. The four men are very hungry, so they agree to Big Kiku's deal. Big Kiku says that the tattoos must be in place tonight, but the cassava root will not be available until the following morning.*

*You learn that Big Kiku hates some of these men for betraying him to his enemies. You suspect that he will cheat and betray some of them. Thus this is the perfect opportunity for you to see first hand how Big Kiku wields his power. The cards below have information about the fates of the four men. Each card represents one man. One side of the card tells whether or not the man went through with the facial tattoo that evening and the other side of the card tells whether or not Big Kiku gave that man Cassava root the next day." (Cosmides, 1989)*

The models that are presented as representing this problem are therefore as follows:

| | |
|---|---|
| *[eating cassava]* | *tattoo* |
| ┐ *eating Cassava* | *[┐ tattoo]* |

*...* *(Johnson-Laird and Byrne, 1991)*

### 2.5.4 Theorists can Change the content of the cards by labelling them with negations.

Cheng and Holyoak (1985) ran an experiment and presented subjects with the following materials that included negations.

"*Suppose you are an authority checking whether or not people are obeying certain regulations. The regulations have the general form 'If one is to take action A then one must first satisfy precondition P.' In other words, in order to be permitted to do A one must first have fulfilled prerequisite P. The cards below contain information on four people: one side of the card indicates whether or not a person has taken action A, the other indicates whether or not the same individual has fulfilled precondition P. In order to check that a certain regulation is being followed, which of the cards below would you turn over? Turn over only those you need to check to be sure." This is followed by four cards, each stating one of the four possible cases: "has taken action*

> *A", "has not taken action A", "has fulfilled precondition P" and "has not fulfilled precondition P*". (p. 403)

Here we find a different type of directionality that compares the meaning of two sets the first is "*A versus Not A*" and the second "*A versus B*" where B is to denote the second possibility thereby negating the first. In the given experiment we have the Not A explicitly stated so it does not have to be inferred from the text as is usually the case in the classical task. If we regard meaning as existing in space we find that the orthogonality of opposite meanings is stated with the Not A possibility and not clearly implied with the classical task are the second letter could also be C, D, E,..etc. Therefore, the reason for better performance could be the explicit negative as a result of what it implies. If there would be a way to imply opposites that are equally orthogonal, it would be interesting to see if the NotQ selections would also rise and this is indeed attempted here.

### 2.5.5 Theorists can Change the task to allow subjects to envisage all the alternatives explicitly.

An example of this case, is when the four possible cards are: P, Q, NotP and NotQ and the subjects are given only two cards, Q and NotQ. The cards removed as the cards that subjects seem to do well with, selecting P and hardly ever selecting NotP. Oddly enough, subjects seem to make the NotQ selection with an extremely high percentage. This task is called the 'reduced array' selection task (Wason and Green, 1984)

Other examples that Johnson-Laird and Byrne (1991) placed under this category are questions that require subjects to verbalise their thought patterns and therefore be led to consider all possible alternatives.

The reduced array task indicates a strong effect of the explicit existence of P on a card because when it is present selections of Q are high while when it is absent selections of NotQ are higher. Since the rule subjects consider is *If P then Q* then it seems intuitive to observe that subjects follow it to arrive at Q when P exists and to doubt the existence of Q when P does not exist.

The theory of Mental Models is a truly admirable, powerful theory that unfortunately does not add much to what Wason (1966, 1969) showed to be inherent to the original task as it simply fortifies the assumptions that if the semantics lead subjects to believe that a violation is likely, then they are more likely to select the NotQ card.

## 2.6 Relevance Theory versus Competence Theory

Now after an avalanche of experiments in the task, of which only a sample has been reviewed within this thesis, some theories emerged to explain behaviour in the task. This section will ponder Relevance Theory versus Information Gain as two representatives that can be compared to each other.

First of all, Sperber et al. (1995) introduced Relevance Theory by relating the design of sentences to a theory of the pragmatics of natural language. They started by highlighting some of the misunderstandings that are possible within the selection task. One is that the rule is general in that it applies to all comparable situations within the domain of the cards in front of the subject and subjects are expected to understand it as such. This is opposite to understanding it as a particular rule that applies to a particular card, making that one card sufficient to prove the rule. The second possible misunderstanding is the narrow versus wide range interpretation of cards. A narrow range would be just the four possible values on the faces of the cards in the question as in just the values A,K, 4,7 while a wide range of possible values may include all vowels, all consonants versus all even numbers or all odd numbers. Another issue that was raised is that of implicit versus explicit content that leads subjects to what is the desired outcome of the task. An explicit instruction would be: *"Did Big Kiku get away with cheating any of these four men? Indicate only those card(s) you definitely need to turn over to see if Big Kiku has broken his word to any of these four men" (Cosmides, 1989,p.265)*. The authors indicate that subject overconfidence in pragmatic processes and relevance issues is what causes them to fail to select the correct cards. On the other hand, subjects succeed when the pragmatics of the version and relevance result in logically correct selections.

Another extremely important issue that was raised by Sperber et al. (1995) is that of how a rule of the form *IF P THEN Q* can be falsified. The authors explain that such a rule can be falsified in two ways that are logically equivalent but not computationally equivalent. The first method is to follow the rule in the order of the premises so if P is true, then a subject has to check whether or not Q is true to check if the rule is false. The second method is to take the implication that P and NotQ will be false if the rule is true. Notice that the first method is dynamic in the sense that the rule is "*followed*" from P to Q while the second method is static in assuming a conjunction of the two premises would falsify the rule. This particular issue will be investigated through several experiments that will "*direct*" subjects towards one or the other interpretations.

The authors also indicate that there is a large difference with respect to implicit versus explicit negatives of each other with respect to semantics. They explain that judging a positive statement to be false appears to be easier than judging an explicit negative as a true statement. This brings up the issue of the difficulty with negatives. One of the direct consequences of this is that saying that a feature is not present is not computationally equivalent to giving an example of a feature that implies the negation of another. In the second case, all subjects do is to judge false a positive statement that seems to be cognitively easier. This matter will also be reviewed in detail in this work only to show that giving just another example is not sufficient to make mental reasoning any easier.

Relevance Theory is introduced as being about what information represents to an individual. Relevance is described as being positively correlated with the cognitive ***effect*** resulting from processing the information and negatively correlated with the processing ***effort*** required for processing a piece of information. The first cognitive principle of relevance is that human cognitive processes are aimed at processing the most relevant information in the most relevant way. The communicative principle of relevance is that every utterance conveys a presumption of its own relevance.

They indicate that the *P-and-NotQ* condition is a complex feature while P and Q are simple features. Therefore, in order to make it possible for subjects to select the *P-and-NotQ* cards then this complex feature must be easier to reflect on than the feature of *P-and-Q*. Take for example a context where volunteers are required to take care of a large number of children. The rule would be "*If a volunteer is male, then he is married*". The subjects would be informed that men don't want to take care of children. This information would cast a doubt on the presence of present bachelors, therefore they are led to test that condition which is the NotQ condition.

If we consider these claims, we find that the theory claims that the abstract task has no relevance relation between the premises and this should explain behaviour in that task. However, it should be clear in the recipe of an easy selection task that it builds on highlighting the possibility of the rule being broken, except here it is done through relevance which is a way of implying the importance of semantics in the task. Consequently, it does not seem to offer any new information on the difficulty faced by subjects when tackling negatives nor on the influence of the background rule. It does however, indicate that this relevance is not restricted to being cued into taking roles or perspectives.

It should be clear to the reader that the authors have shown a greater understanding of the influence of semantics in the task and represented this through relevance. The difficulties expressed with negatives and the selection of P and NotQ do indeed exist.

Perhaps an ideal complementing theory to the above is that of Information Gain as proposed by Oaksford and Chater (1994). The critical issue here is that this theory offers an 'inductive' explanation of a task that has to date been regarded as purely 'deductive'. They suggest that subjects choose cards such that they provide the greatest information gain in deciding on what cards to select. They start by assigning probabilities to the various possible selections that are either as part of an independent model or dependent model. The independent model is where selections of q are completely independent of the selections of p. From there the $E(I_g)$ which is the expected information gain is calculated as follows. Their model shows the informative behaviour of all four cards as follows:

1. *p card: Is informative insofar as P(q) is low. It is largely independent of P(p).*
2. *q card: Is informative when P(p) and P(q) are both small.*
3. *not-q card: Is informative to the extent that P(p) is large. It is independent of P(q).*
4. *not-p card: Is not informative.(Oaksford & Chater, 1994)*

Additionally, we are informed that the changes in the probabilities $P(M_I)$ rescales the $E(I_g)$ values but does not change the order of selections of the four cards. We are also informed that $E[I_g(\text{not-p})]$ is always zero as we cannot deduce anything if the antecedent is not true. Additionally, when P(p) and P(q) are small then $E[I_g(q)]$ is greater than $E[I_g(\text{not-q})]$. Another thing is that at all other values of P(p) and P(q), either $E[I_g(\text{not-q})]$ is greater than $E[I_g(q)]$ or these values are undefined. There is an assumption that by default P(p) and P(q) are rare and this implies low probabilities for both resulting in the order of card selections as *P > Q > NotQ> NotP* for the abstract versions of the task.

They inform us that for "thematic" questions the task is different because subjects are not given the task of *rule testing*, instead they are given the task of checking *rule use*. With rule testing the model calculates information gain while with rule use the expected utilities are calculated as subjects use the rules to maximise expected utility. They assume a small cost for turning any card, while a utility is gained by the information offered by turning that card and seeing the other side. They also inform us that subjects associate particular utilities with the adopted perspective for the rule, so a positive utility assigned to rule violations is larger than the cost of turning over the card. This in effect would alter the selections of the cards to the order P > NotQ > Q > NotP.

If one is to compare the two theories, the first by Sperber et al. (1995) is concerned with the semantic implication of the premises and how 'dependent' they are on each other with respect to what they would imply as part as background knowledge. However, what they do explain is the same issue of raising the possibility of a falsifying instance, except that here it is done through semantically relating the premises to each other. One should not jump to the conclusion that this relation is useless as it will be revisited in Chapter 4. The second theory comes from a purely statistical starting point to arrive at the same conclusions. Its principles assume by default a 'rarity' of probabilities and that the abstract task is completely different from the "thematic" version which makes the two different tasks result in different orders.

## 2.7 Conclusion: Perspectives and Descartes

Perhaps at this point of time, one should take a philosophical stop to ponder the path of research to date. Ever since researchers first uncovered the initial "*error*" there has been a frantic search for "*successful*" materials. Therefore, a clear neglect of the classical task is evident in the abstract task while a wealth of tests was conserved for the "*thematic*" versions, which were gradually more specifically described into necessitating pragmatic schemas with a goal or rationale and then the necessity of having them as a social interchange with the possibility of cheating. Figure 2.1 shows a meta diagram of the research path followed to date.

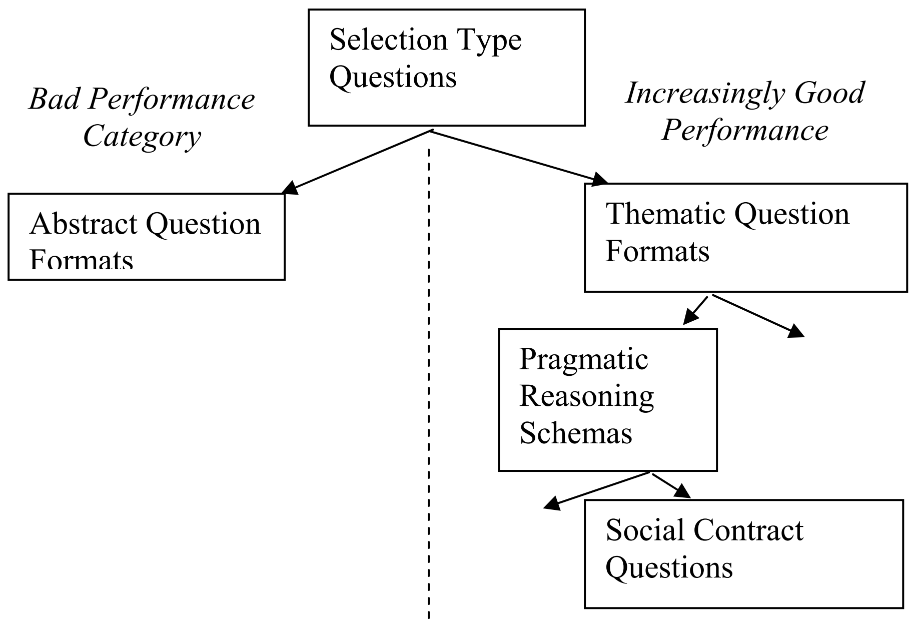



Figure 2.1: The Division of Materials into two groups one more heavily researched than the other.

With so little work done to test versions of the original task or abstract versions that are similar, it is not surprising that researchers "*assume*" that they are aware of what is going on with that version of the task, to enable them to concentrate on testing the more interesting right hand side of that figure. Just to show how misleading this assumption is looking at table 2.1 will show the differences in percentages of correct selections of P and NotQ for various abstract task benchmarks used as a measure of comparison with "*thematic*" type problems.

| **Sources of Abstract Test** | **Number of Subjects** | **Percentage selection of P and NotQ** |
|---|---|---|
| Wason 1968+1969 | 128 | 4% |
| Wason and Shapiro 1971 | 16 | 12.5% |
| **Cheng and Holyoak 1985** | **22** | **19%** |
| **Cosmides 1989** | **24** | **25%** |
| Stenning and van Lambalgen 2004 | 108 | 3.7% |

Table 2.1: A comparison of benchmark figures for "*correct*" selections in the abstract task

More experienced researchers though will note that with the higher number of subjects, as in the first line, the effect becomes more stable. However, clearly no one has ever thought of comparing the rest of the data to identify how it can be 'informative'. We can probe the size of the difference by taking a Chi test of the data shown above as a whole, while assuming no particular set as the expected outcome in order to just estimate the degree of difference between them. A chi test of the whole set of values reveals an unsurprising Chi value of 20.713 with $p < 0.0004$. The most distant studies from Wason's original are those of Cheng and Holyoak with a Chi value of 7.263 with $p < 0.007$ and Cosmides with a Chi value of 13.145 and $p < 0.0003$. Since a significant difference is informative in general psychological research, it is odd that no one noticed that the original task seems to have this problem that it is extremely difficult to replicate to statistically acceptable percentages.

If the benchmark base of the data can differ significantly from one replication to another, then this can only mean that there are indeed effects that can vary subject behaviour in the basic task. It is possible, therefore, that these effects have gone unnoticed as no one has considered to date the need to formally compare performance in the classical or binary form of the classical task across studies. It is also possible that the assumption that nothing can cause improvement to these tasks has also been taken for granted without looking at why Cosmides' (1989) abstract results are so high considering the question was within a booklet amongst the other problems, nor look at Cheng and Holyoak's (1985) results where there was a recorded improvement close to significance.

Another issue that seems to call for attention at this point is that of perspectives where a congruent alignment with the direction of the rule seems to produce a higher degree of accuracy than a non-congruent alignment as is shown by the work of Gigerenzer and Hug (1992). For example, if a subject is cued into an employee's perspective and given the rule "*If an employee works on the weekend, then that person gets a day off during the week*" then that subject is likely to do better than a subject cued into an employer's perspective and given the same rule. These results clearly show the influence of directionality on the task. And if that is not enough, Cosmides (1989) attempted a form of "*private exchange*" that brings to mind Descartes' beliefs. In his book *Rules for the Direction of the Mind* (Descartes, 1954)[4] he indicated a desire for truth that could only be satisfied through two mental operations: intuition and deduction. Intuition represents our understanding of self-evident principles such as the axioms of geometry. Deduction, on

[4] Translation copy.

the other hand, represents orderly reasoning or inference from self-evident propositions as all of geometry is reasoned in strict order through deduction from its self- evident axioms and postulates. This may be the basis of logical deduction as it exists today, to start off with a rule and some given premises and to deduce or infer from them what they obtain.

However, this deduction must be in accordance with a specific order of matters. "***The chief secret of the method is to arrange all facts into a deductive logical system***". Whether one agrees with Descartes' general philosophical views or not, one cannot deny that this statement clearly describes the process which subjects of the task are "*supposed*" to follow in order to arrive at a conclusion that is correct and true.

> *"I perceive so clearly that there exist no certain marks by which the state of waking can ever be distinguished from sleep, that I feel greatly astonished; and in amazement I almost persuade myself that I am now dreaming."(Descartes, 1641, Meditation I)*

This is early on during his first day of meditation when dreams become similar to reality. He starts by questioning everything, even reality, and arrives after a long deliberation during his second meditation to the famous ***cogito ergo sum***, or ***I think, therefore I exist.*** He arrived at this conclusion following from the assumption that he is being deceived and so long as that is occurring, he can never be nothing (Descartes, 1641). From here, we find a seed that implies that perhaps "*egoism*" or the influence of the "*I*" may affect subject performance in this task. This exactly what Cosmides (1989) describes as a "*private exchange*", even though her theory is evolutionary rather than logical.

Sperber et al. (1995) also show that a rule can be falsified by two methods. The first is through "*following*" the rule if P is true, then one can go on to check whether or not Q is true, if NotQ is true then the rule is false. The second is to check the negative which is that P and NotQ is true and if so then the rule is false. So the first implies a dynamic reasoning approach while the second implies a static reasoning approach. Additionally, they also indicate that another influencing factor could be the implicit versus explicit negatives. An example is that *white* is an implicit negative of *black* while *not white* is an explicit negative of *white*. They indicate that the implicit negatives should make processing easier than explicit negatives.

Still regarding the issue of negatives, the Oaksford and Chater model (1994) also predicts that subjects comprehend the rule with a negative antecedent and positive consequent more slowly than the one with a negative antecedent and negative consequent and the one with a positive antecedent and negative consequent.

To sum up, we have various indicators here to factors that seem to strongly affect performance in the task, the most important of which is the difference between the abstract and “thematic” versions of the task. The chapter started by reviewing how the problem was found and classified as an “*error*”. This was followed by a search of “*successful*” materials which resulted in subjects selecting the competent response. Then an overview of Socratic tutoring showed that this is a neglected avenue that may indeed prove itself to be a powerful tool that was initially utilized by Wason and Johnson-Laird (1970). Then the path went on to discover the effects of rule familiarity, matching bias, suppositional reasoning, pragmatic schemas and social contracts. All this work was done in the light of a basic assumption that the competent answer is indeed the “*correct*” one. Clearly, the path does not seem to be going towards any solution to the problem with the classical abstract format with regards to explaining subject behaviour although there are excellent indicators in the work done to date of what might be the culprit. Consequently, the following chapter presents a much needed new way of looking at the problem by discussing what Stenning and van Lambalgen (2002) contributed to this division.

# Chapter 3

# A Novel Perspective

*"As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality."(Einstein)*

## 3.1 Introduction

Einstein shows us that the beautiful world of symbols like P is very far from the reality that the words reflect. There exists a challenge represented in the following two rules:

*"If there is a tiger in the forest, then there is a hunter in the forest.*

*If there is a tiger in the forest, then there is a worm in the forest."*

It is easy to notice that the Q value in the first sentence is very different semantically from Q's value in the second. In the first we are talking about a hunter that means to remove the danger posed by the tiger being in the forest, while in the second the worm has nothing to do with the tiger and simply co-exists with him.

Classical logic treats both in the same way, as it is blind to the semantics behind this Q value, while in real life things are completely different. Although logic tempts logicians with a beautiful perfection, as Einstein notes above, it does not conform to the world we are actually living in. The solution therefore can only lie in a different type of logic that is more sensitive to the semantics of the premises so that it can help us make sense of the world. The logic that seems appropriate for the above example is deontic logic, where in the first rule we find that the existence of the hunter in that forest is something that 'ought to' happen due to the danger that would exist if it does not.

Stenning and van Lambalgen (1999, 2001, 2002, 2004) have detected this type of anomaly and offered us a new perspective from which we could examine subject behaviour in the selection task without being isolated from semantics via a classical logic restriction. However, before going into the details of their contribution, we should first look into the indicators that emerged from prior studies that make this conclusion the most reasonable to arrive at.

## 3.2 Indicators

The first and most important indicator is that researchers in the task did recognise and admit that the "*thematic*", "*pragmatic*" and "*social contract*" tasks are in effect deontic. Cheng and Holyoak (1985) inform us of the following:

> "*The permission schema, in contrast, contains no context free symbols such as p and q above. Instead, the inference patterns include as components the concepts of possibility, necessity, an action to be taken, and a precondition to be satisfied. (The deontic concepts of possibility and necessity are typically expressed in English by the modals of can and must, respectively, and various synonyms, such as may and is required to.)"(p.396)*

From Cosmides (1989) we have:

> "*The experiments reported here must have primarily tested for the presence of the 'look for cheaters' procedure; however, their design also involves, and to some extent tests, some features of the interpretive component's implicit inference procedures.*
>
> *The most important example of this implicit inference is the subject's interpretation of the deontic relations in the rule itself. In a situation of social exchange, in order to be entitled to receive a benefit, one is obligated to provide a benefit, usually at some cost to oneself. These deontic concepts should be inferred by the interpretive component, even when they are not explicitly mentioned in the rule. The deontic operator 'may', indicating entitlement, should be assigned to the benefit clause, and the deontic operator 'must', indicating obligation, should be assigned to the cost clause, no matter what their position in the social contract rule. Thus, social contract theory predicts that when the interpretive component recognizes a term as a cost or benefit, it will cause the appropriate deontic operator to be assigned.*"

These indicators also happen to represent some of the most successful contexts in the task and the authors indicate that these theories are linked to deontic reasoning or semantics. Other indicators stem from the lack of a clear distribution conformance of the subject selections between the different attempts of the abstract form of the selection task as was shown in the conclusion section of Chapter 2. The study will go into more detail in Chapter 5 to compare behaviour in the individual selections of PQ, P, Q, etc., in order to isolate similarities from differences. It is truly odd that no one has delved into this type of comparison previously just as a post analysis of the behaviour in the abstract task itself. Perhaps in the midst of all that attention, researchers overlooked this type of comparison even though it is highly indicative that the different populations that were tested could in fact respond differently according to their background information, level of knowledge and the like that may affect performance in the abstract task and not only the "*thematic*" version alone. If anything, it indicates that the confusion that results from the abstract task is far from being predictable. Although a large percentage of subjects select PQ, the other divisions do not seem to directly conform to any particular division that is statistically reliable. This leads one to the conclusion that it should be intuitive that when a task is deontic, then subjects and researchers have a clear line of communication between them that is not subject to any noise or confusion, while in the case of the abstract task, subjects do not conform to specific results because there is only confusion in the air.

The third indicator in the list is to simply raise the question: Is it reasonable to assume that selecting P and NotQ are indeed the correct answers to the task? This issue has been raised in Chapter Two to show that assuming that these are the correct answers is based on Popper's (1963) philosophy and unfortunately theorists have been falling into the same error as their subjects by ***not*** seeking falsifying evidence to their theories. Clearly, the case with the "*thematic*" type questions is that they are easy to solve because they are not like the abstract tasks while the theories are more like the abstract tasks in their distance from the actual examples of the questions. This raises a serious question about what it is within the "thematic" questions that makes them easier while it is absent from all forms of abstract representations.

## 3.3 What is Deontic Logic

All this talk about the effect of deontic semantics on the task must surely make the reader curious as to what is implied by it and what deontic logic is defined as. Well, as much as one may search, it is the most obvious of definitions that are extremely hard to find. It seems that everyone seems to take this type of logic for granted such that everyone forgot to define it and presumed that a definition exists somewhere. The longest definition of this logic exists in the Stanford Encyclopaedia of Philosophy. It states that Mally (1926) presented the first formal system of deontic logic and that it was faulty and Menger (1939) regarded it as unacceptable. The Encyclopaedia provides no formal definition of what deontic logic is. Other entries include the one for Mirriam Webster's dictionary (2003) that only states that deontology is "*the theory or study of moral obligation*".

Clearly there is a lack where definitions are concerned, so perhaps we can turn to the psychologists who introduce this type of logic for a clearer definition. This we can take from Manktelow and Over (1991)

> *"Deontic thinking takes place when we consider what we may (or are allowed to) do, or what we ought to (must or should) do, rather than what was, is, or will actually be the case. So basic forms of deontic thought are those concerned with permissions and obligations. Cheng and Holyoak (1985) are responsible for introducing deontic considerations into research on conditional reasoning, a development which has invigorated the field. They propose that there are schemas acquired from experience, and stored in long-term memory, for deontic reasoning about what they take to be permission rules…"(p.88)*

Manktelow and Over (1987) p.231

> *"In this chapter we are interested in one extremely important type of deontic reasoning, which takes place when people try to find out which actions they ought to perform or may perform. This type of reasoning has traditionally, in philosophy, been called 'practical reasoning' and is distinguished from 'theoretical reasoning', which has the object of trying to discover, or to describe correctly, objective matters of fact. It is sometimes said that the difference between these two is that between trying to infer what one should (or may) do as opposed to trying to infer what one should (or may) believe. The latter does not have to be 'theoretical' in the scientific sense and could be directed towards ordinary facts which are highly relevant to practical questions about what one should or should not do. For example, facts about what is healthy or unhealthy. Practical reasoning depends, in part, on some*

> *degree of theoretical reasoning, but goes beyond it to conclusions about actions…."*

Therefore deontic logic came from the moral obligation or wanting to perform an obligation and we are also informed that it relates to 'practical reasoning'. In spite of what seems to be a clear definition, it is not precise and perhaps one reason for this lack is that the concept was born in the vague world of philosophy. However, this lack does not render it unusable because it has indeed found itself in the very well defined world of logic. Mally (1926) stumbled in his first attempt to formulate it, but that did not last, as the Standard Deontic Logic was defined by Follesdal and Hilpinen (1971) and along with it many other types of logic emerged including preference-based, defeasible and temporal deontic logics.

However, it is not advisable at this stage to jump into the logical ocean without learning how to swim. Therefore, we will start with a clear definition as to what would constitute a deontic task to us and what would on the other hand constitute a descriptive task. We will follow that with some deductions based on the definition of each.

**Definition 1:** **Deontic Rule.** A rule shall be considered a deontic rule if the existence of a negative instance does not cast any doubt on the existence of the rule merely on the conformity of the case. Instead, a negative instance would be regarded as a violation of the rule.

**Example:** An example from Bahrain's roads could be: "*If one is to drive, then that person must have a valid driving licence.*"

**Implications:** Consequently, if one is found driving without a driving licence and that is not unheard of, then it does not cast doubt on the rule.

**Definition 2:** **Descriptive Rule.** A rule shall be considered a descriptive rule if the existence of a negative instance casts doubt on the truth of the rule.

**Example:** An example could be "*If there is an A on one side, then there is a 4 on the other side.*"

**Implications:** It does not associate with any moral "*obligations*" or "*permissions*" and therefore the mere appearance of an A on one side, and a 7 on the other side is enough to cast doubt on such a rule: it may be an exception, but it may also be a counter example.

The basic assumptions are that deontic implication can either be expressed through the rule itself as shown above, or through the context, or through the background knowledge of the question. The context can express moral obligations or permissions by informing subjects that they are "*an authority checking…*" (Cheng and Holyoak, 1985). Background information is assumed in the case of the drinking age rule for example. This in fact is generally the case, as the surface form of the sentence out of context is frequently ambiguous as to whether it will be interpreted deontically or descriptively.

Consequently we find a nice clear division between two types of tasks: the descriptive tasks and the deontic tasks. This analysis to the sceptical reader may not seem sufficiently strong to adopt this perspective. To simply state that "*deontic*" rules are those that do not cease to be in force if a negative instance arises does not give any explanation of why this particular format succeeds while the descriptive one fails. Yet again we seem to be dividing materials into two parts: "*successful*" and "*unsuccessful*" types of questions.

This is why Stenning and van Lambalgen (2004) conducted a detailed study of the descriptive task, in order to identify the paths of thought that subjects follow and to understand why subject behaviour is the way it is.

## 3.4 Paths of Thought in the Abstract/Descriptive task

Stenning and van Lambalgen (2004) carried out an investigation into the various paths of thought that subjects could have followed in order to arrive at their different conclusions in selecting PQ, PQNotQ, P, etc. The scheme they followed is that of Socratic tutoring in order to guide subjects along the path in a way not much different from that followed by Wason and Johnson-Laird (Wason and Johnson-Laird, 1970).

They did, however use a two rule task of the form shown below:

*Rule 1: If there is a U on one side, then there is an 8 on the other side.*

*Rule 2: If there is an I on one side, then there is an 8 on the other side.(Stenning and van Lambalgen, 2004)*

Subjects are informed that one of these rules is true while the other is false. This makes the competent answer to this question to check the card with 3 because if this card has a U on the other side, then Rule 1 is automatically rendered false and Rule 2 is rendered true because one of them must be true. If, on the other hand, the other side of the 3 has an I then Rule 2 is automatically rendered false and Rule 1 rendered true.

One of the main advantages of this rule format is that it backgrounds concerns about exceptions as some subjects, as will be shown in section 3.4.1, presume that some of the rules presented may still hold even though an exception may exist for that rule. The two rule task is further discussed in section 3.5.2.

A careful analysis of the two rules presented showed that the rules are of the structures: If P then Q, and If NotP then Q. This would eliminate the possibility that subjects relate judging the truth of the card to their trust of the person performing the experiment because there are two options to select from, rather than say the rule is true or false.

Stenning and van Lambalgen (2004) started to draw a map of possibilities based on semantic implications and these were then observed in tutorial Socratic dialogue. Subjects were asked to verbalise their reasoning to help isolate five distinct problems on which we may focus:

1. The problem of exceptions.
2. Truth of the rule and 'truth of the card'
3. The logic of "true"
4. The logic of "false"
5. The cards as sample

The problems presented here emerged during Socratic Dialogues with the aim of understanding subject reasoning. Consequently, they are inherent to the task and any manipulation that will be attempted in the forthcoming chapters that results in altering subject behaviour would probably achieve that purpose through amplifying or resolving the problems described here. It is for this reason that a thorough review is essential to aid one in recognising the causes and effects as well recognising if any other problems exist.

### 3.4.1 The Problem of Exceptions

Some subjects seem to accept the possibility of having an exception to the rule while the rule still holds. It could be a coincidence or not that, in English, exceptions are evident with respect to the rules of grammar. Consequently, the status of the rule being in English may influence or cause this type of behaviour by showing that it is natural to expect rules to be robust to exceptions.

Stenning and van Lambalgen (2004) indicate that if a counterexample is not found as sufficient evidence that the rule is false, then it would seriously complicate matters. If subjects assume a deontic mental reasoning perspective we find that

> *"In this paper we study the violability of norms, and in particular how to proceed once a norm has been violated. Clearly this issue is of great practical relevance, because in most applications norms are violated frequently. In the fine print of a contract it is therefore usually stipulated what has to be done if a term in the contract is violated. For example, if the delivery time is overdue the responsible agent might be obliged to pay the extra transport and warehousing costs that result from the delay. If the violation is not too serious, or was not intended by the violating party, the contracting parties usually do not want to consider this as a breach of contracts, but simply as a disruption in the execution of the contract that has to be repaired."(van der Torre, 1999)*

I realise that the authors intend their isolation to be of descriptive tasks, so imposing this paragraph that shows that this behaviour is evident in deontic contexts may be somewhat surprising. There is no central contradiction because in deontics we find that what a person focuses on is what "*ought to*" happen while in the case of simply assuming exceptions, there is no indication of what the existence of an exception means. For example, it may exist because the rule applies to another set of data and this particular instance is not part of the domain.

The issue of subject confusion should be obvious to any researcher in the field and will be investigated in detail through the experiments in Chapter 5.

### 3.4.2 Truth of the rule and 'truth of the card'

Here we find a clear problem with alignment of the mapping between the premises of the rule and the sides of the cards. Such an ambiguity may cause subjects to behave as follows:

"*Subject 10:*

> *E: If you found an 8 on this card [I], what would it say?*
>
> *S: It would say that rule two is true, and if the two cannot be true then rule one is wrong ...*

> *(subject turns 8).*
>
> *E: OK so it's got an I on the back, what does that mean?*
>
> *S: It means that rule two is true.*
>
> *E: Are you sure?*
>
> *S: I'm just thinking* ***whether they are exclusive****, yes because if there is an I then there is an 8. Yes, yes it must be that." (Stenning and van Lambalgen, 2004)*

There is a confusion here either to whether the two rules are "*exclusive*", in that if one is confirmed then the other cannot be confirmed without risking having both true, or whether the four cards are exclusive. The researchers indicate that here subjects 'transfer' the truth of the rule in the individual case of the card to that of the rule. They are therefore assuming that each card defines a domain of its own.

However, one may also note that what should be the case is that the two cannot be falsified at the same time, and not that the two cannot be confirmed at the same time as subjects may assume to be the case. Therefore, a confusion in reasoning alignment may cause subjects to apply the principle, that should be applied to falsification, erroneously to confirmation.

### 3.4.3 The Logic of "true"

The logic of "*true*" occurs when some subjects insist that "*not false*" is not the same as "*true*". A sample dialogue is presented where a subject saw that rule 1 is false and rule 2 is not false, but that subject could not arrive to the conclusion that rule 2 is true. The reader may recall that if one rule is false then the other must be true by virtue of the problem design.

This behaviour, we are told, may persist in the selection task as turning P and NotQ suffices to know that the rule is not false but it does not prove the rule to be true. By contrast in 'deontic' cases we are told that subjects can never prove a rule such as the following to be true, therefore they just try to ensure it is not violated: *" If you want to drink alcohol on these premises, you have to be over 18".* Perhaps this interpretation may cause subjects to select all cards to be extra sure of their conclusions.

### 3.4.4 The Logic of "false"

The term "*strong falsity*" was defined when a subject equates a false conditional "*If P then NotQ*" with the false occurrence of the rule If *P then Q*". Note that this conclusion was obtained through a subject that selected P and Q. An example of this case is presented as follows:

*Subject 26 (Standard selection task):*

> *"E: So you're saying that if the statement is true, then the number [on the back of A] will be 4 ... What would happen if the statement were false?*
>
> *S: Then it would a number other than 4." (Stenning and van Lambalgen, 2004)*

Here we are told that subjects may regard counterexamples as governed by a rule and in this case the cards P and Q are sufficient to show that the rule is not false, possibly because they have a bi-conditional interpretation of the rule in addition to strong falsity. Strong falsity alone seems to guide subjects to select "*either P or Q*". Note that this is not the only suggestion that will be given to guide subjects to select P alone or Q alone, so the paths proposed here are simply a guide of possible explanations not a clear mapping from subject choices to interpretations.

### 3.4.5 The Cards as Sample

This we are informed is the more 'natural' way of interpreting the rule so that it applies to the full domain of letters and numbers rather than just to the four cards involved in the question. This generalisation would either lead subjects to take the cards as a sample from cards that can many other values and numbers, or to resort to probabilities to arrive at a conclusion.

In this case, we find the subject trying to judge using some form of probability function that does not seem very different from that proposed in the Oaksford and Chater (1994) model of information gain.

### 3.4.6 A Directionality in the Paths?

If there is a directional aspect to thought, then it should be present in all seasons and interpretations however subtle the effect may seem. We can start with the problem of exceptions, and find that an exception is overlooked when the subject is intent on going on with the reasoning process unhindered by the issue that arose. Here, what we have is a determination by the subject to follow along in the same path of thought in spite of possible threats to the assumptions made. Gebauer and Laming (1997) argue that constant anaphora and biconditional interpretations are persistently held. They presented the cards of the selection task six times to each subject and found that subjects chose the same cards as each other. They turned the cards subjects chose in each of the tests. However, they did not provide any feedback to students that would alert their attention to consequents of choices or causes. Although this evidence is not sufficient to defend the assumption that reasoning in this task is not dynamic, it does seem to show a tendency to follow that path if there is no external interference with the path of thought selected.

The anaphora that requires relating the premises of the rule to the two sides of the cards is none other than an alignment problem. Some Subjects restrict themselves to constant anaphora or reading the rule as saying "*if A is on the face of the card, then 4 is on the back of the card.*" This indicates that dealing with the meaning of the words "*one side*" and "*other side*", relating that to the precedent and consequent of the rules and in turn to what is displayed on the face and back of the cards seem to cause a significant alignment problem to subjects. One should notice that in each of these cases there are two options to each of the possibilities. If one is regarded as a "*not*" of the other, as is expected by logic, then this reasoning should be subject to the problems classically encountered when expecting subjects to interpret "*not false*" into true. What this means is that when subjects assign the face of the card to the letter A and the back of the card to a number, then they are checking perhaps 4. If they then wish to look at the reverse such that the side that has the letter is on the back, then the number will appear on the front. In the second case, we have the A not on the face of the card, and at the same time it may not be an A. These possibilities of having negatives in what is unknown seem to be caused by the anaphora and explained by the difficulty of handling negatives.

In the two rule task, the reader may recall that the two rules cannot both be true or both be false which places them both at opposite sides of the spectrum of truth/falsity and although it backgrounds some issues, it does not resolve this one.

The logic of false follows where subjects apply the "*Not*" to the Q part by regarding counter examples as governed by a rule, yet again, we find an organised world that has both positive and negative examples clearly outlined. Here subjects do not wonder about the falsity of the precedent, thereby keeping it constant while doubting what may emerge as a consequent. This behaviour shows a classic sensitivity to the direction of the rule.

Last but not least, we find the last student continuing to talk about verification even when that student turns the card with 3, does not seem to have any implications of falsification. Here we find a contrast between the tasks themselves that are assumed by students, where some may follow the path of verifying a rule, while others follow the path of seeking to falsify the rule. Yet again these two seem to have contrasting pragmatic goals and consequently different "*ways*" of looking at the problem.

Kirby (1994) ran a number of experiments where he tested the effect of the set size giving P for example a value from 1 to 1000 and a rule that says if a card has a number from 1 to 1000 on one side, then it has a + on the other side. He noticed that the selections of P,Q and the selections of NotP, NotQ were inversely affected such that the former decreased as the set size rose, while the latter increased as the set size rose. This suggests an alignment between the similar values of truth or falsity. This perhaps can be informative in suggesting a form of directionality in the selections. Note that the Q and NotQ values that he used as + and – are implicit negatives of each other.

These indicators imply that the semantics of the task may be capable of influencing subject interpretations and in turn influencing their behaviour in the task. Although directionality will be further discussed in Chapter 4, it is worth mentioning here that guiding subjects away from one erroneous assumption does not necessitate that all these subjects will arrive at the same conclusion. Instead, what occurs should be similar to highlighting one or more danger zones that they should avoid causing them to continue to fluctuate between other problems that continue to exist.

## 3.5 Predictions and Experiment

From the paths of thought isolated we find a main conclusion that the abstract or descriptive task is definitely ambiguous. It is glaringly obvious that this great diversity of interpretations could only be possible if there is no clear communication between the theorists who put the questions and the subjects who attempt to perform the task.

Stenning and van Lambalgen (2004) ran an experiment to test if subjects do indeed follow the paths of thought that emerged from the Socratic Tutoring sessions. Since the interpretations subjects seem to make are guided by the semantics of the question, then manipulations must imply a semantic setting that is different from the baseline. They attempted five different conditions that are as follows:

1. The classical task to be used as a ***benchmark***
2. The two rule task: ***Expected outcome more accurate responses of NotQ***.
3. Exploring conjunctive rules: ***Prediction that subjects will not make competence response to not turn any cards and will give it a deontic interpretation***.
4. Contingency instructions: ***Prediction of an increased selection of NotQ***.
5. Judging truthfulness of an independent source: ***Prediction is that this would raise accuracy levels.***

The first condition replicates the distribution reported in Wason (1968) while conditions 2 and 3 alter the logical format of the rule with the aim of "backgrounding" erroneous assumptions. Conditions 4 and 5 keep the rule as is, while addition extra textual clarification to the questions.

### 3.5.1 Investigating Different Logical Formats of the Rule

The two rule task has the same first paragraph as with the classical task while the second paragraph is shown below:

> *"..Also below there appear two rules. One rule is true of all the cards, the other isn't. Your task is to decide which cards (if any) you must turn in order to decide which rule holds. Don't turn unnecessary cards. Tick the cards you want to turn.*
>
> ***Rule 1:*** *If there is a vowel on one side, then there is an even number on the other side.*
>
> ***Rule 2:*** *If there is a consonant on one side, then there is an even number on the other side."* (Stenning and van Lambalgen, 2004, p.31)

This is the only format where the selection of P is not part of the "*expected*" selections, instead the correct response is Not Q. The authors inform us that this format should background concerns about the significance of exceptions as well as other issues.

The second task that alters the rule format has a conjunctive rather than the classical "if then" structure, and the following rule was tested:

> ***Rule:*** *There are vowels on one side and even numbers on the other side. (Stenning and van Lambalgen, 2004, p.34)*

This is a statement that for all the cards, there is a vowel on one side and an even number on the other. So the existence of cards with a consonant and an odd number are enough to know the fact that the rule is false without turning a single card. However, the authors predict that subjects will turn cards reflecting that the rule does in fact have a deontic force that implies that they would check it as a rule as a result of the context it is embedded in.

### 3.5.2 Investigating the Effect of Extra Textual Clarification

The contingency formulation of the question aims at removing any contingency relations as may be understood by subjects.

> *"…Also below there appears a rule. Your task is to decide which of these four cards you must turn (if any) in order to decide if the rule is true.* ***Assume that you have to decide whether to turn each card before you get any information from any of the turns you choose to make****. Don't turn unnecessary cards. Tick the cards you want to turn."(Stenning and van Lambalgen, 2004, p.32)*

The authors inform us that the prediction is for better performance with this format.

The second format of this type attempts to avoid confusing subjects into believing that they are judging the experimenter's credibility when testing the rule, and the following paragraph is added:

> *"…Also below there appears a rule* ***put forward by an unreliable source****. Your task is to decide which cards (if any) you must turn in order to decide* ***if the unreliable source is lying****. Don't turn unnecessary cards. Tick the cards you want to turn." (Stenning and van Lambalgen, 2004, p. 33)*

### 3.5.3 Results from Stenning and van Lambalgen (2004)

The frequencies of subject choices as provided by the authors on p.35 are shown below in table 3.1:

| Condition | PQ | Q | P | **P¬Q** | ¬Q | ¬PQ | PQ¬Q | ¬P¬Q | All | None | Misc. | total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Classical | 56 | 7 | 8 | **4** | 3 | 7 | 1 | 2 | 9 | 8 | 5 | 108 |
| Truthfulness | 39 | 6 | 9 | **14** | 0 | 7 | 3 | 6 | 8 | 15 | 5 | 112 |
| 2-rule | 8 | 8 | 2 | 1 | **9** | 2 | 1 | 0 | 0 | 2 | 4 | 37 |
| Contingency | 15 | 0 | 3 | **8** | 1 | 6 | 4 | 8 | 3 | 0 | 3 | 51 |
| Conjunctive | 21 | 1 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | **1** | 2 | 30 |

Table 3.1: Stenning and van Lambalgen's (2004) results

Researchers would be primarily interested in the bold values, which show the correct responses according to classical logic are given. If we set aside the 2-rule task because it has a different form, we find that the best performance appeared of course when the truthfulness of the source is placed in doubt with Contingency in second place. It may be hard for some to imagine how informing subjects that they must judge all cards without thinking they can turn the other side, would result in such a strong positive effect, yet this is true. Perhaps what this instruction does is to point subjects to focus on the "*fronts*" of the cards. Kirby (1994) noticed an issue regarding the "*front*s" of cards where he had a higher selection of the P and NotP cards when they came from smaller sets. One of the possible conclusions he raised was that of attention bias. This suggests that a subject may utilize the card's ability to draw attention as a heuristic device for assessing the relevance of the card to the condition (Evans, 1983).

If we look at the percentage selections of the four cards in the Classical and Conjunctive formats as shown in Table 3.2 we find a redistribution of selections. A particularly clear difference exists in the increased selection of negative values of NotP and particularly NotQ which goes from 15% to 48%.

| Condition | P | Not P | Q | **Not Q** |
|---|---|---|---|---|
| Classical | 72% | 14% | 73% | **15%** |
| Contingency | 65% | 34% | 55% | **48%** |

Table 3.2: Percentage selections of each card

We can also go on to ponder around the other results shown above, but before we ponder values as they are, we shall convert the table to percentages in table 3.3 as the nominal values are confusing for a comparison of this sort.

| Condition | PQ | Q | P | **P¬Q** | ¬Q | ¬PQ | PQ¬Q | ¬P¬Q | All | None | Misc. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Classical | 52 | 6 | 7 | **4** | 3 | 6 | 1 | 2 | 8 | 7 | 5 |
| Truthfulness | 35 | 5 | 8 | **13** | 0 | 6 | 3 | 5 | 7 | 13 | 4 |
| 2-rule | 22 | 22 | 5 | 3 | **24** | 5 | 3 | 0 | 0 | 5 | 11 |
| Contingency | 29 | 0 | 6 | **16** | 2 | 12 | 8 | 16 | 6 | 0 | 6 |
| Conjunctive | 45 | 3 | 13 | 10 | 3 | 0 | 0 | 1 | 0 | **13** | 12 |

Table 3.3: Percentage selections out of the totals

One of the clearest differences in the PQ selection column is the low percentage of selections for the 2-rule task. Perhaps the form of the problem that is represented as two rules "*If P then Q*" and "*If NotP then Q*" while expecting subjects to find which one of the two holds affects performance here. In the case of this question, we find that both P and NotP are explicitly mentioned, so if we apply a form of pattern matching technique to the rules a selection of P, NotP and Q may be more likely. This is not one of the categories displayed above, which is a strong indication that selections of this combination are not high. A more careful look at the 2-rule task shows that while PQ selections are low, there are relatively higher selections of Q alone and of Not Q alone. The task therefore, seems to guide subjects towards focusing on the consequent of the rules because the two possibilities are explicitly displayed in the antecedent of the rule.

This variation of the task, therefore, seems to implicitly background issues about the nature of the relationship between the P and Q premises that appear in the rule. We find that subjects who in other variations place them together to compose their selection of PQ, in this task start to regard them as independent entities, such that P can be ignored when concentrating on whether to select Q or NotQ. While this variation guides 46% of the students to selecting one of these two alternatives, they do not select both together. On the other hand, in other variations up to 52% of subjects actually do select PQ together. Two reasons may cause this separation: either it is the format of the problem through giving a selection of one of the two rules because the truth of one proves the falsity of the other; or it is the difficulty of processing negatives along with positive values, and here again we have the same truth versus falsity comparison.

## 3.6 Are Deontic Questions the only ones capable of "*successful*" behaviour?

If one is to follow Popper's (1963) philosophy then any theory should be prepared to stand up to sincere falsification attempts. In the case under study here, the assumption that deontic questions are those that resulted in successful behaviour is a form of post analysis of research data. One may be tempted to assume that this type of question which is currently described as a "*thematic*" question is the only type capable of eliciting the competent responses of P and NotQ.

To first test the claim, one must check that deontic questions that express an ideal actually do succeed in eliciting good performance. The work of Cheng and Holyoak (1985), Cheng et al. (1986), Cosmides (1989) and Gigerenzer and Hug (1992) support the claim that the successful questions they used are all deontic. However, in order to "*prove*" the claim, an exhaustive testing of all possible deontic conditionals has to be tested. Clearly this is an impossible task to perform at least within the limits of a researcher's lifetime.

However, one can look at some specific cases that may be regarded by some as deontic, and reported to have failed to produce "*successful*" competent responses. The first and clearest is when a rule has the word "*must*" or is in a form like "*If A is on one side, then 4 must be on the other side.*" Here, some may be misled into presuming this is sufficient to allow the rule to be regarded as a deontic rule. However, if the reader refers back to the definitions given in section 3.3, it should not be difficult to realise that even with the word "*must*" the rule is still as fragile as other descriptive/abstract rules in its category. Consequently, this form does not pose any threat to the assumption that deontic questions are those that to date elicit more of the competent answers.

Another possible challenge is posed by the transportation rule. This is a strange case because it only succeeded once when tested by Wason and Shapiro (1971) and failed in later attempts (Griggs & Cox, 1982; Pollard, 1981; Manktelow & Evans, 1979; Gilhooly and Falconer, 1974). Researchers attempted to find reasons for the failure by finding differences in the procedures followed in the replications, like whether or not the four days of the week are written on the same cards as the possibilities, written on separate cards, or not at all. Other issues that arose were that the original task was run on a per student basis and each was given instructions alone. Possible other issues that may cause the difference may emerge from location differences between the postal rule for example "*If a letter is sealed then it must carry a 20-cent stamp*" when tested by Cheng and Holyoak (1985) in Michigan as compared to Hong Kong. The people of Hong Kong are well acquainted with this rule while those in Michigan are not. Consequently, even in the no rationale version of the problem, the subjects of Hong Kong did extremely well in the task. By analogy, one may assume that the subjects that took the transportation problem had a form of background knowledge that would in turn influence their behaviour such that they would assume there to be an ideal method of transport that should be followed.

Subjects in that particular experiment are from the University of London which perhaps accidentally is the starting point of exactly two railway lines going north and two large motorways going north. Additionally, one may also note that each of the railway lines and each of the roads are either to the east or west of the country, therefore making it more appropriate for one to take them to go to either Leeds or Manchester. This information, coupled with perhaps some social information of the transport service available as well as the availability of cars, may readily impose some form of ideal situation that does not apply when the same test is run elsewhere. An example could be of a tourist someplace in the world asking a hotel concierge about the best way to travel to another city. One possible answer is: "*There are two ways to go there: by train or by car. I would personally take the train because traffic in these parts is just horrible and would get on your nerves.*" Another possible answer is: "*There are two ways to go there: by train or by car. I would personally drive to it because the road is wonderfully scenic and traffic is extremely light.*" Now, if we speed up the clock and increase the number of cars on the same scenic road, we find the second situation turning into the first. The point is that the "*ideal*" that is imposed by background information onto transport is dynamic, because the number of cars on the road increases at a high rate and train services get better or worse with time depending on management. Consequently, the level of comfort that is sought by the traveller may differ thus affecting the situation. If the two means become comparable, then it is expected to find that subjects behave in this task as if it is descriptive. However, if one means of travel is much more interesting and comfortable than the other, and the rule contains the uncomfortable one, then it should not be surprising to find that subjects make more selections of the competent response.

So there does not seem to be any evidence of the existence of any deontic question that failed to elicit adequate competent response behaviour. Therefore, we can look at the other group of tasks which contains the classical task and all the tasks that seem to not imply any ideal to find that they elicited very low accuracies as is also replicated in the Stenning and van Lambalgen (2004) study. Now, if there is any descriptive task that is capable of eliciting "*successful*" competent response behaviour, then the deontic type of materials would not explain behaviour and instead would only represent part of the materials that exhibit the traits of "*successful*" materials.

Before we jump to the conclusion that no descriptive task is capable of attaining a high degree of accuracy, one may consider a task tested by Kirby (1994) that tests the size of the set of P versus the set of NotP. He found that the rule "*If a card has a 0 on one side, then it has a + on the other side*" elicited around 40% accuracy with 18 subjects out of 45 selecting the P, NotQ cards. Subjects were then shown cards with the values 879, 0, +, -. However, when Kirby (1994) alters that rule to "*If a card has a number from 1 to 1000 on one side, then it has a + on the other side*" while keeping the same card choices, then selections of P, NotQ drop to 20%. However, this did not replicate well with the other experiments attempted as performance was roughly the same with respect to competent response selection.

The experiments did show a trend that links selections of NotQ to the set size of P such that the larger the P set is, the more likely subjects are to select the NotQ card. With the 2-rule task we also find a higher selection of NotQ with both the 2-rule task and the contingency conditions.

Altering the set size of P increases the number of cases that have to be tested for verification assuming it covers a domain that is larger than the four displayed cards. If subjects are concerned with the whole domain of 1000 numbers, then it would be mentally taxing for them to attempt to verify the rule and falsifying it when only two values of the consequent exist represented in + and – seems more attractive. In the 2-rule task both P and NotP are part of a rule that may be chosen as the one which is true, leaving the NotQ selection being the only unknown that is not displayed. In this case, it seems to be a balance between what is known and what is unknown. So the sets here have attributes of whether they are explicitly displayed or not rather than the size of the set being of concern. In the third case, what is resolved is a dependency between card choices and how one selection may influence the other. When a subject selects a card, the subject must know that the other side will remain hidden when having to select the next card. Here the two sets are formed by what is known and what is not known and each card is considered individually.

Clearly, the domains of P, Q and the cards, whether part of a larger domain or each in a domain of its own, do affect performance in this task, but how they do is not at all clear. One thing is evident though and is of primary concern here, that while the domains do affect performance they do not seem to present themselves as candidate features for the "*successful*" abstract/descriptive task that is desired as a falsification attempt to the assumption that all "*successful*" questions are deontic. Whether or not such a task exists still eludes researchers, so the jury is still out until such a task actually shows itself and proves to be widely replicable, and the theory stands without any serious falsification attempts.

## 3.7 Anaphora and Directionality

The work of Stenning and van Lambalgen (2001, 2004) highlights the existence of an anaphor between what is on the face of the card and what is on the other side. They show us a subject that explicitly changes the direction of the implication in the following dialogue:

*Subject 12 [experiments 1a,1b,1c]:*

> *"E: The first rule says if there is a vowel on the face of the card, so what we mean by face is the bit you can see, then there is an even number on the back of the card, so that's the bit you can't see. So which cards would you turn over to check the rule.*
>
> *S: Well, I just thought 4, but then it doesn't necessarily say that if there is a 4 that there is a vowel underneath, So the A.*
>
> *E: For this one it's the reverse, so it says if there is a vowel on the back, so the bit you can't see, there is an even number on the face; so in this sense which ones would you pick?*
>
> *S: [Subject ticks 4] This one.*
>
> *E: So why wouldn't you pick any of the other cards?*
>
> *S: Because it says that if there is an even number on the face, then there is a vowel, so it would have to be one of those [referring to the numbers].*
>
> *:*
>
> *E: [This rule] says if there is a vowel on one side of the card, either face or back, then there is an even number on the other side, either face or back.*
>
> *S: I would pick that one [the A] and that one [the 4].*
>
> *E: So, why?*
>
> *S: Because it would show me that if I turned that [pointing to the 4] over and there was an A then the 4 is true, so I would turn it over. Oh, I don't know.* ***This is confusing me now because I know it goes only one way.***
>
> *(Stenning and van Lambalgen, 2001)*

The problem of dealing with the rule that goes one way, while the face and back of the cards can be exchanged, keeps coming up again and again through the literature. Stenning and van Lambalgen (2001) indicate that there is an interference effect that is even evident in subjects that are able to give the normative responses that occurs when they look at the cards in both orders, letter up or letter down.

> *"So an 'interference' explanation for the choice of the p,q card would run like this. Suppose subjects decompose the intended variable anaphora reading of 'one side – other side' into 'face/back' and 'back/face', and then proceed to reverse the direction of the implication in the latter case. This would lead to the transition from*
>
> *If there is a vowel on one side, then an even number is on the other side*
>
> *via*
>
> *If there is a vowel on the face, then an even number is on the back*
> *and*
> *If there is a vowel on the back, then an even number is on the face*
>
> *to*
>
> *If there is a vowel on the face, then an even number is on the back*
> *and*
> *If there is an even number on the face, then a vowel is on the back*
>
> *What speaks in favour of this analysis is that about one third of our subjects consider the K/4 card to be irrelevant, whereas 4/K is taken to falsify,.. a surprising fact, which is however entirely consistent with the analysis proposed here. What speaks against it, however, is that some subjects who give the normative answer for the intended reading of the rule, reverse the arrow in case of the 'back/face' anaphora."(p. 21)*

By contrast, in deontic questions, this anaphora does not seem to exist. The context of the questions seems to make the possibility of a negative instance existing easier to recall. For example, in the rule "*If the form says entering on one side, then the other side includes Cholera on its list of diseases*." It is clear to anyone reading this rule that to ensure that people had their inoculation against Cholera is something to be desired and equally clear that those who have not had their inoculation may attempt to enter the country without it. What this implies is that the relationship that is held between entering with or without an inoculation is not a random one, as there is an 'ideal' to not wish people to be ill as a result of allowing someone to break this rule. Consequently, what relates the first premise in the rule to the second is more than just a random relationship that has no consequences whether the rule is held or broken.

On the other hand, in abstract or descriptive questions, subjects have to recall the relationship because it bears no importance to them and while doing so they need to align it with the precedent and consequent of the rule as well as the background rule that is provided. This requirement for alignment and the errors shown so clearly by Stenning and van Lambalgen (2001, 2004) present a picture of what may be a background effect that influences behaviour. One of the problems of course is the logic of "*true*" where subjects are unable to equate "*not false*" with "*true*". The problem with processing negatives has also to be clarified throughout the review and if this problem is compared to the one with anaphora we find the same patterns emerging; namely a problem with evaluating what part of the rule maps onto what side of the cards and how to keep the background rule in mind during the process. In short, what some subjects feel is that this is similar to a "*juggling*" game where they are not allowed to let any of the balls fall onto the ground.

So where is the directionality? The problem in anaphora is caused by a lack in subject abilities to map the rule onto the front/back of the cards. When presented by a rule such as "*If A is on one side, then 4 is on the other side*" and cards that show A, K, 4, 7, they directly apply the rule to A and search what is behind A, and this indeed is chosen by most subjects. The problem arises when the other cards are under consideration, because if subjects wish to verify the rule, they would wonder which card had an A on the back to check if 4 is on the front and usually the one to be selected is 4. If on the other hand, they are following a falsification track, they would attempt to select cards that would cause the rule to be false and in this case would select the expected A and 7 because they would search for the card that has an A on the back and not a 4 on the front. Clearly the competent answer is the one that is guided by a pragmatic falsification goal and to date no experiment has managed to guide subjects to think along this path.

In other words, the deontic/descriptive division, shows that subjects are strongly influenced by the method or approach they use to reason about the problem. If an ideal exists, then there is a reason to think of what "*ought not*" be done and consequently they think along the falsification track and make the competent selections of P, NotQ. If, on the other hand, the task is not deontic, then they are unable to diverge from the standard verification path that causes them to get confused.

This is sufficient to indicate a strong effect of "*directionality*" but rather than stop there we can consider the rule: "*If there is no water in your radiator, your engine will overheat immediately*"(p.25). We are informed that *The presence of 'your' is what makes* *the interpretation more likely to be hypothetical; the antecedent need not ever be true for 'your' car*"(Stenning and van Lambalgen, 2001,p.25). What would occur if it is true or can be implied as true? Here we have a difference made between *event conditionals* that describe events that we are told are situated in real time, and *hypothetical conditions*. So the utilization of these words such as "*I*" and "*you*" are capable of taking conditionals from those sensitive to real time events to those that live in hypothetical worlds. These hypothetical worlds are 'ideal' worlds where there are things that 'ought' and 'ought not' happen.

Chapter 4 will present an introduction to the different types of directionality and then impose a temporal sequence that ensures that the event referred to by the antecedent always occurs before the event referred to by the consequent through a conveyor setting.

Chapter 5 will utilise information packaging to identify and attempt to alter subject focus and imposed distinctions onto the task. A negation task which investigate accepted assumptions while altering the hierarchical structure of the referents produces striking results.

Chapter 6 will analyse surface co-occurrence features of rules to identify if a distinction exists at that level as well as attempt to figure out its justification.

# Chapter 4

# Selection Task Reasoning in Time

Subjects who reason in the selection task are given a question from which they are expected to extract assumptions that would in turn be utilised to help them arrive at their conclusion which will eventually appear as responses. This is not a point of controversy as such, because all theories of reasoning have the same assumptions but vary in the degree to which the process is dynamic. Explanations of what causes the various conclusions subjects arrive at vary widely assuming different types of sequential progression. However, the existence of these types of "*order*" does not inform us whether or not subject reasoning can be influenced by introducing sequences in time.

Consequently, this chapter will start by considering some of the possible sequences that may exist in the reasoning task. The experiments presented here will then introduce a uni-directional temporal restriction and this will be imposed onto the task in order to transform the semantic anaphora of "*one side/other side*" into a sequence with three possible locations instead of two.

## 4.1 Sequences in Reasoning

The aim of this section is to identify some of the possible sequences in reasoning that may imply a sequence in time. These offer themselves as indicators of a possible interaction between reasoning and a temporal context in the question. The aim is to identify if such a semantic context is a candidate that can influence reasoning by "*inviting*" interpretations that are not likely in other contexts.

Several types of sequences will therefore be introduced followed by a section on possible interactions with an imposed timeline.

### 4.1.1 Logical Directionality

This is the directionality of logical deduction, which is mainly "*ordered*" with respect to logical deduction. For example, if we have the rule *If A, then B* and the premise A is given, we can logically deduce B and it is exemplified by the following set of statements:

A (is true),

**If A, then B**

⇨ B

This type of directionality is represented by the path of proof followed by mapping the true A to the antecedent and then arriving at the consequent. One of the main characteristics of this type of directionality is that B is deduced from A even if the rule is of the form *B only if A*.

Whenever several possible paths are possible, then the consequences of alternative actions are left open to the subject. Pragmatic reasoning schemas as introduced by Cheng and Holyoak (1985) exhibit what is meant by reasoning in order to achieve a "*rationale*" or a goal. The rationale is not represented by the deduced B as in the simple form shown here but is instead shown through a "*goal*". An example is the case of a permission context as in the rule "*If the form says 'ENTERING' on one side, then the other side includes Cholera among its list of diseases.*" The authors claim that it qualifies as a permission context because of the rationale of "*This is to ensure that the entering passengers are protected against the disease*"(p.401, Cheng & Holyoak, 1985). The idea is that this form of rationale is what causes the permission schema to be invoked and that in turn is a set of four production rules that are followed in order to guide subjects to make the deductions necessary to arrive at their conclusions.

It is important to note that this type of rules also presents an example of 'deontic' type rules where subjects may think in the context of "*if this is what ought to happen, it does not mean that it must necessarily happen*". Here we find that logical deduction leads subjects to assume the role of falsifying the Cholera rule and consequently arrive at the competent logical solution.

In order to identify rule features that affect performance such as whether or not they are goal driven, a comparative analysis was conducted here of 79 experimental runs.[5] Complete results of this analysis are included in Appendix A as they proved to be a valuable source of information as based on several studies (Cheng and Holyoak, 1985; Cheng et al., 1986; Wason and Shapiro, 1971; Manktelow and Evans, 1979; and Gigerenzer and Hug (1992). The data are only presented by comparison of performance percentages because no sufficient baseline replications existed in the various studies to allow any formal meta analysis of the experiments to be conducted. The variable baseline, made it impossible to convert any of the statistical scores like the F score or Chi value into a common measure of difference that could be mapped onto a unified scale. Consequently, although the data are informal, it may be informative by providing indicators of the factors that affect performance.

The data were first broken up into several groups each dedicated to a different range of performance percentages. The groups of rules that resulted were as follows:

**Group A: Super-performer (percentage correct >= 80%):** This group included 15 runs and all fell into two main categories, either as social contracts with up to 95% accuracy rate or as permission or regulation schemas when a rationale is given. It may be worth mentioning that the social contract rules are descriptions assigned to these tasks by the experimenters and this has in some cases been subject to controversy. Most of these tasks involve explicitly informing subjects that they should look for a cheater or that this is obviously implied. In fact a further analysis shows that all of these 15 have a deontic context, a deontic rule, or both.

**Group B: Moderate-performers (percentage correct < 80% and >= 56%):** This group included 29 runs with only one "*abstract*" case. This "*abstract*" rule is the rule in Cheng and Holyoak (1985) "*If one is to take action 'A' then one must satisfy precondition 'P'*". This case is deontic because it implies an obligation but it is also abstract because it is a general form rather than an example of a situation. However, it is not similar to the abstract forms of the rule that are usually descriptive. In this case, 22 of the 29 runs have a deontic context, a deontic rule or both, while the rest are purely descriptive.

[5] An experimental run is whenever an experimenter gives subjects a particular question at a particular time and elicits responses, so the same rules run by different researchers or at different times qualify to be different experimental runs.

**Group C: Weak-performers (percentage correct < 56% and > 25%):** This group included 19 runs of which 8 are contingency training subjects and 8 are obligation training subjects of Cheng et al. (1986). Most of these rules merely describe correlation rules that relate seemingly independent antecedents to consequents and arrive at this group as a result of the training subjects are exposed to. Here, 11 of the 19 runs have a deontic context, a deontic rule or both.

**Group D: Marginal-performers (percentage correct < 25%):** This group included 16 runs most of which had no clear relation between the two propositions other than mere correlations, and subjects are not exposed to any form of training. In this group only 2 runs have a deontic context or a deontic rule.

First of all, the existence of a deontic setting within these runs and with decreasing percentages of 100%, 76%, 58% and 12.5% as one goes from Group A through to Group D shows that the competent response selections margin seems to be correlated with the frequency of deontic rules in that particular group. Once a rule is interpreted descriptively not deontically, then subjects seem to follow a diverse set of possible logical paths as shown by Stenning and van Lambalgen (2004). This is echoed by the data given above to show that deontic runs are much more frequent in the higher group levels than they are in the lower achieving ones as shown in figure 4.1.

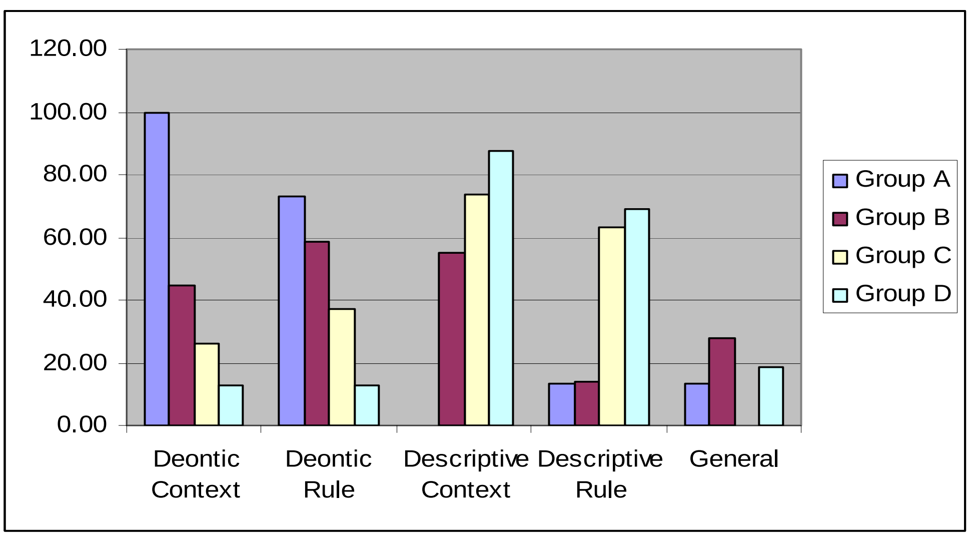



Figure 4.1: The percentage of Deontic questions or contexts shown as per Group[6]

[6] The "General" category includes all rules that cannot be easily classified into any of the categories shown. Rather than imposing judgement and forcing classification these few rules were placed into this category.

However, this order of logical deduction, although it seems to form the basis of assessing how "*correct*" student responses are by comparing them to the competent logical response, does ignore the existence of other types of directionality that may interfere with subject performance when the problem is not precisely described. In the case of deontic questions, subjects exhibit behaviour that reflects that they fully understand the experimenter's intentions. By contrast, with the abstract/descriptive task, they exhibit a confusion that may be caused by interferences of different possible paths. One possible indicator of the existence of different forms of directionality was presented by Cheng and Holyoak, (1985) who tested the "*If..then*" structure versus the "o*nly if*" structure and found that subjects reason significantly more accurately with the first structure than with the second structure. They also gave subjects a rephrase task and the significant difference between the two structures was evident again to a significant level when assessing how many subjects produced competent rephrased statements. The "*only if*" structure differs from the "*if then*" structure in that the conclusion comes ahead of the condition in the former, so subjects are likely to learn what they can deduce before they learn what is necessary for that deduction.

### 4.1.2 The Directionality of Knowledge Acquisition

This type of directionality represents the order in which the subject acquires knowledge, whether this is dictated by the experimenter or the subject's logical deduction process. Here we find that the knowledge presented to subjects in this task is given in the form of two rules; a foreground rule and a background rule so the order of acquisition may differ from one subject to the next depending on the order in which they make the associations between the cards and the clauses in the rule. It starts with the task presented to the subject and a rule similar to the one shown in Figure 4.2 assuming cards shown contain A, B, 7, 11.

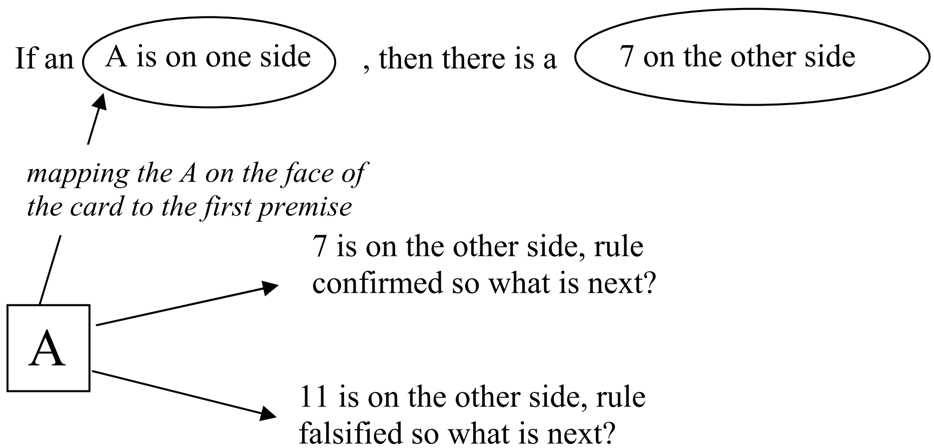



Figure 4.2: A graphical representation of one of the paths that may be followed.

Subjects may start by considering the first card and utilising it to deduce the influence of what is on the other side of the card on the task at hand. If the other side has a 7, then the rule is confirmed and this assumption may be carried forward to influence their perspective of the second card as they consider it as a possible answer. If, on the other hand, the other side has an 11, then the rule is falsified by that finding and they may carry this assumption while considering the next possible card. The anaphora exists between the syntactic relationship between "*one side*" and "*other side*" that affects the semantic relations that necessitate a letter to occupy one while a number occupies the other. Therefore, the contents of the face of the card can be mapped onto either the antecedent or the consequent clauses. This causes a mapping from the two sides of the cards to the antecedent and consequent clauses that is variable according to the card. This clearly is a source of confusion that may lead to two basic different interpretations (Stenning, 2002).

Since the anaphora is variable, comprehending its full scope of application is complex. One possible way of interpretation may be a constant anaphoric interpretation where the face of the card is always mapped to the antecedent premise and it should lead subjects to only select the A card because then it would be the only one relevant (Stenning, 2002). Another possibility is to interpret the rule as a bi-conditional with constant anaphora causing the competent solution to be A and 4, which is incidentally the highest selected subject solution to this task.

Perhaps it is worth considering here that there may be interference between the different kinds of directionality. While the first interpretation results from applying the distinction between the antecedent and consequent to the face and back of the card, the second interpretation results from applying the reversibility of the cards to the antecedent and consequent of the rule, causing it to be bi-conditional. This difference between the two bits of knowledge offered through the question itself, has different directional biases that may offer an explanation as to why subjects would attempt to force an alignment between them in order to arrive at their selections.

In the case of constant anaphoric interpretation we find that the "*face*" of the card is mapped to the antecedent perhaps because it captures attention before the "*back*" of the card. This exhibits a directional interpretation that is restricted by this relationship that is introduced by the subject from the first premise to the second, and from the face of the card to the back.

The variable anaphora make the problem more complex so one way of simplifying it could be to erroneously interpret it as a bi-conditional. Perhaps the asymmetry of the If/then construct clashes with the front/back inter-changeability of the card so the inter-changeability is applied to the rule turning it into a bi-conditional. With a bi-conditional, the mental process that takes a subject across the rule from P to Q has to make a round trip bringing the subject back from Q to P, causing subjects to select the P and Q cards as the optimal choices to check the rule.

The contingency variant of the task (Stenning and van Lambalgen, 2004), for example, resulted in a major change in subject behaviour, increasing selections of NotQ to 48% as compared to 15% in the baseline task. The only difference between this variant and the classical task is that subjects are explicitly informed the following; "*Assume that you have to decide whether to turn each card before you get any information from any of the turns you choose to make*". What this instruction does is eliminate any dependency issues between card choices, such that each card should be selected as if it is the first that is considered in the group. What this essentially implies is that the path of reasoning that takes a subject through the group of cards under consideration carries with it a number of dependencies from one to the other. When the statement is included subjects become aware that this is not desired and are therefore not prone to the interpretations that allow these dependences to exist. This supports the conjecture that the difference in directional features may cause dependencies to be carried from the rule to the cards or vice versa.

### 4.1.3 Tensed Directionality

Tenses impose an event timeline that starts in the past and ends in the future. The standard tense order for the classical selection task is present-present in each of the premises. One can refer again to the analysis shown in Appendix A to investigate whether any effect of tense directionality exists and whether or not this would affect performance. A careful comparative partial analysis of this data showed that some of the experimenters did not always conform to tense restrictions. The study covered a careful analysis of each time a rule is tested experimentally, and this is what is described by a run. This analysis will be referred to at various points in the thesis.

Results showed that of the 79 runs, 9 have a present-present perfect tense ordering, 9 present-future ordering and 3 past-present ordering and the rest a simple present-present ordering. Details are shown in table 4.1

| Tenses | Total number of rules | Number of rules scoring above 50% | Average Percentage Performance of the rules | Average Percentage Performance of all other rules |
|---|---|---|---|---|
| Past-Present | 3 | 1 | 54.33% | 53.3% |
| Present-Present Perfect | 9 | 8 | **65.67%** | **51.8%** |
| Present-Future | 9 | 7 | **62.3%** | **52.2%** |
| Present-Present | 58 | 29 | 50.00% | 62.62% |

Table 4.1: A classification of the rules according to different tense combinations

If these numbers are tested through a Chi Square distribution to see if any variation exists between the group that has different tenses in the premises and the group that has a strict present-present tense, the results show a Chi value of 4.314 and a $p < 0.0378$ with a Yates corrected p-value of $< 0.06$ which is bordering on significance. The effect of the deontic/descriptive classification has already been shown to correlate with better performance so the tense analysis here may interfere with those results to produce these differences.

Sometimes a deontic task that has a present/present tense semantically implies that the consequent state occurs as a result of an action that took place at an earlier point in time than the antecedent as in Cosmides (1989) and Gigerenzer and Hug (1992) as shown in the rule: "*If a man eats cassava root then he must have a tattoo on his face.*" In this example one can easily see that for the tattoo to be present on the face, it must have been obtained at a point in time prior to which the eating process should be occurring, thereby implying a temporal distance on the timeline that is not reflected by a change in tenses.[7]

Since the task itself is highly sensitive to semantic implications the tensed directionality is able to impose an effect of its own on the effectiveness of the rule which is difficult to detect, perhaps because as in the cases shown above the timeline differences may be implied rather than clearly shown. A more specific study dedicated to study this effect within a single population was run of the Gigerenzer and Hug (1992) experiments. A meta form analysis was run of the data provided by the authors and results showed a Standardized Mean Difference of –0.1396 which is hardly sufficient to claim a difference between rules with ordered tenses and those who have both verbs in the present tense. The same set of 24 rules show no differences with respect to the order of the tenses in the two premises.

However, the lack of sufficient information for analysis does not imply that this form of tensed directionality does not exist nor can it deny an effect on subject behaviour. The given percentages above do in a sense show a trend where percentage competent responses seem to rise whenever different tenses are employed in the same rule.

Abstract tasks pose a difficulty when one wishes to utilise such tensed ordering or any timeline distancing to emerge because they seem to impose a static presentation of the question. One cannot say '*If there was an A on one side of the card, there will appear a 4 on the other side of the card*' and expect it to be semantically equivalent to the classical abstract question. It seems therefore safe to assume that current data does not show that tenses play a vital role in abstract rule interpretation. This does not imply that they are neutral, rather it implies that it is extremely difficult at this point in time to arrive at solid conclusions based on the experiments covered in the analysis.

### 4.1.4 Cause and Effect or Symptom and Cause

These two types of directionality would either be in the form ***If*** *cause* ***then*** *effect*, or ***if*** *symptom* ***then*** *cause*. This type of rule does not exist in abundance in the literature as it only seems to exist in the Cheng et al. experiment to test the rule "*If two objects carry like electrical charges, then they will repel each other*." It is a lonely rule so it is hard to

[7] Another interpretation that may be implied is that if a man eats Cassava root then that man will "have a tattoo" placed on his face at a later point in time so here too we have the "order" implied between the two events. However, Cosmides (1989) explains that the process takes place one night before the men are offered food so this interpretation is not likely in the case of her experiment.

compare it to others with respect to performance issues to identify the effect of causality on subject behaviour. It results in a 34% accuracy or P, NotQ choices which seems high when we consider that the baseline for that batch of experiments is a selection task replication with 19% competent responses which is itself quite high. Wason's results by contrast show a percentage selection of 4% that is quite replicable, so the lack of replication in this case hinders further analysis.

Judgement of whether a rule reflects a causal link is based on a clear causal link directly displayed between the precedent and consequent as in the case of rules of physics as presented in the rule "*If two objects carry electrical charges, they will repel each other*".[8] This classification is extremely strict in the sense that their relationship is not a causality implied by the If structure and exists independent of the rule form. The latter type forms almost all If instructions and includes many different classifications which are not relevant to this study.

Due to the lack of proper testing of this semantic relation in the analysed data, it is difficult to arrive at any solid conclusions on whether or not it has any effect on behaviour.

### 4.1.5 Discourse Directionality

Here the issue of central concern is that of the order in which a subject goes through the reasoning process from one conclusion to the next. This overlaps with logical directionality, but because subjects deduce B from A when confronted with a rule of the form If A then B, and overlaps with Knowledge Acquisition in that B only if A, it is not equivalent to the above.

An example of this type of directionality in the case of the selection task is when a subject examines the rule in order, so if the rule has the form B only if A, and the premise A is given, then the subject can follow a path of thought that reads the rule in the order it was given, and then map the given A to the rule to deduce B. The sequence of thought here is restricted to the sequence given in the rule as it is first guided by it. Additionally, the sequence of thought is also restricted by that of logical deduction because B is deduced even in the given rule order. There are no assumptions made here that following

---

[8] The type of cause/effect relationship described here is very different from that imposed by the If structure as defined by Fillenbaum (1978) because the target is that the relation is part of the semantic relations between the premises.

the two different orders as part of this directionality is not likely to cause confusion, and all that is being said is that it is possible to incorporate the two under a larger umbrella.

> *"The consequentialism question needs to be separated into two distinct components: (a) do people think through the consequences of choices before deciding and (b) does any such analysis determine the choices made? On the evidence of the inspection times analysis presented here and of protocol analysis reported elsewhere it appears that subjects do think about the consequences of choices on the selection task, but think far more about the cards they end up choosing than those which they reject." (p. 238, Evans, 1996)*

Evans presents strong evidence that what seems to be happening is that subjects think a great deal about the cards they choose which are mostly the P and Q selections, but they hardly think at all about the ones they do not choose. Perhaps the rule, as it sits in front of subjects, gets far more attention than the cards, and therefore in a sense "*guides*" subjects along its path which does indeed contain the P and Q cards, especially since the classical abstract question hardly contains the word 'not' which may be taken to denote the NotP and NotQ cards that exist, which is what Evans called "*pattern matching*" behaviour. Another possibility is that the interpretations that lead to the selections of the P and Q cards warrant much more consideration than those of the others because they are more readily adopted. In this case the consequentialism exists in the process of going from the basic interpretation to the conclusion that causes card selections such that the simplest process is being selected. Therefore, Evans seems to support the notion of the existence of a form of directionality although he does not clarify the sequence implied by his work.

Evans and Newstead (1977) also explain that the reason why selections of P exceed those of Q in premises that are temporally related is due to a mapping of P and Q to events on a timeline. In other words, choices of P increase when P event occurs ahead of Q event, but perhaps Q event could occur before P. Unfortunately, the claim was only made of tasks that are temporally related to each other, so no effort was made to check if the same assumptions extend to the abstract task that only has temporal relations represented in the order of reading the rule or in the point of subject focus as it shifts from the antecedent to the consequent.

One should consider for example what happens in deontic/"thematic" tasks. If one runs a careful examination of the tasks, it should not be hard to identify that in many cases they lead subjects to a contradiction that highlights the NotQ condition either explicitly or implicitly. An example of this is the Cosmides (1989) question shown below:

The rule given to subjects is: *If a man has a tattoo on his face then he eats Cassava root.*

Subjects are given the following information (see section 2.5.3):

> "*You are an anthropologist studying the Kaluame people, a Polynesian people who live in small, warring bands on Maku Island in the Pacific. You are interested in how Kaluame "big men" – chieftans – wield power.*
>
> *"Big Kiku" is a Kaluame big man who is known for his ruthlessness. As a sign of loyalty, he makes his own "subjects" put a tattoo on their face. Members of other Kaluame bands never have facial tattoos. Big Kiku has made so many enemies in other Kaluame, that being caught in another village with a facial tattoo is, quite literally, the kiss of death.*
>
> *Four men from different bands stumble onto Big Kiku's village, starving and desperate. They have been kicked out of their respective villages for various misdeeds, and have come to Big Kiku because they need food badly. Big Kiku offers each of them the following deal:*
>
> *"If you get a tattoo on your face, then I'll give you cassava root."*
>
> *Cassava root is a very sustaining food which Big Kiku's people cultivate. The four men are very hungry, so they agree to Big Kiku's deal. Big Kiku says that the tattoos must be in place tonight, but that the cassava root will not be available until the following morning.*
>
> *You learn that Big Kiku hates some of these men for betraying him to his enemies. You suspect that he will cheat and betray some of them. Thus, this is a perfect opportunity for you to see first hand how Big Kiku wields his power. The cards below have information about the fates of the four men.*"

Here we have what "*ought to*" happen given in the rule, as a deal has been offered and it "*ought to*" be followed. However, Cosmides (1989) also explicitly informs subjects that there is a possibility and motive for Big Kiku to cheat. "*You learn that Big Kiku hates some of these men for betraying him to his enemies*" provides a motive for the cheating while "*You suspect that he will cheat and betray some of them*" gives definite and explicit information that NotQ is highly likely.

This presented example, therefore shows a clear direction that a subject can follow to arrive at a Not Q conclusion. Stenning and van Lambalgen (2001, 2004) have analysed a large number of Socratic Dialogues with subjects and isolated major possible interpretations that would lead most of the subject pool to arrive at their conclusions. The repeated occurrence of the main types implies that it is possible that the set of possible task interpretations seems limited in nature. They have not as yet run a study of what may cause subjects to make one interpretation versus another. But they have run a contingency version where subjects were explicitly informed to judge each card as if it was the first they had considered while not assuming any prior information from other cards.

Since explicitly telling subjects to avoid dependencies between card choices made a big difference in behaviour, the assumption that the cards under consideration are subjected to some sort of "*reasoning order*" is possible. What they did in the this version of their experiment was in fact "*guide*" subjects to avoid the sequence of considering cards one at a time while recalling dependences and replacing it with another path of discourse directionality of assuming each card to be independent of what comes ahead of it, as when starting to consider an independent problem. This implies that dependencies may cause interference amongst different possible paths.

To sum up, several different types of directionalities have been defined as clearly as is possible given current research. The first is logical directionality, which takes a subject from the assumptions to the conclusion with rules that may or may not follow those of classical logic. The second is the directionality of knowledge acquisition, as we all learn, hear, see things in some type of order and this can be made more complex if the presented information includes anaphora as it offers different paths of applying the anaphora to the concepts at hand. Last but not least is the directionality of discourse, which includes the direction followed by the cognitive reasoning a subject adopts as well as the presentation of information. Here, we find that the level of ambiguity and possible emergent different directions rise well above the other two as the directionality of discourse combines them under its umbrella. These differences may possibly cause interference to each of the others driving subjects to apply the restrictions of one part of the given information to the other as was explained in the case of anaphora. However, the occurrence of such interference has not yet been fully investigated.

One may think that all different possible types of directionality have been covered, until one is reminded that when asked if 8 is said before 4 while one counts then does it mean that the direction goes from 8 to 4 or that it is normal to assume that 4 should appear before 8.

## 4.2 Directionality of Thought

The previous sections covered the different interpretations possible in this task and to this point; this one has covered different possible types of directionality, or sequences that may be followed during the reasoning process. However, the sequences or types of directionality that have been discussed are all inherent to the task itself.

Peter Wason (1987) emphasised the view that once subjects follow a path of reasoning that leads to a response other than P, Not Q, then they will never arrive at a point of complete insight even if subjected to Socratic Tutoring. In other words, these subjects are purposely guided as they reason about the question towards the desired responses but they still do not make them.

This process of Socratic Tutoring imposes onto subject reasoning an "*order*" of reasoning by asking them questions about particular cards that subjects may not consider if left to reason on their own. A second approach that imposes order while reasoning about cards is to consider them one at a time in a sequence allowing co-dependencies to emerge. A card therefore is evaluated in terms of all prior information including conjectures made about previously considered cards.

Stenning and van Lambalgen (2004) offered a third path during the contingency version of the abstract problem. All they did was ask subjects to consider feedback from each card as independent of that of any other card in the group of four. Results showed an increase in the competent logical response from 3.64% to 15.69% while P, Q choices are reduced from 50.91% to 29.41%. This clear change in behaviour shows that subjects did indeed assume dependency information from one card to the next as they planned in a sequential format, which is in effect a sequential directed behaviour pattern to thought that may be influenced by imposing a temporal context.

Consequently, this chapter attempts to investigate the effects of imposing a strict temporal sequence that ensures that the event referred to by the antecedent always occurs directly before the event referred to by the consequent. In order to comprehend the origin of the imposed sequence and justification for it, one must consider prior work on the analysis of timelines and events.

## 4.3 Timelines and Events

Evans and Newstead (1977) were the first to emphasise the high percentage of P selections in the selection task and the only explanation it received is that it comes ahead of Q in the rule. However, no solid verification attempts have been made of this explanation nor has any light been shed on the importance of this feature, nor its ability to affect selections.

It is almost impossible to study reasoning directly because the conceptual system is not open to direct external inspection. However, since Popper (1963) described the method of Conjectures and Refutations as one of scientific discovery, and the competent logic response is based on his theory, explanations will be offered as conjectures that await refutations and stand until they appear.

Stenning and van Lambalgen (2004) hypothesise that through Socratic Tutoring sessions subjects resisted accepting that not-false is equivalent to true even if they were guided along that path, and this resistance displays the "*logic of true*" which leads them to **selecting all cards**. They also indicate that when subjects exhibit strong falsity by saying that *Not (If P then Q)* is equivalent to *If P then Not Q*. This may lead them to selecting the option of "*either P or Q*" which is not available amongst possible choices, so **they end up choosing P,Q**. They also show that some subjects assume a constant anaphoric view of the rule which is "*If A is on the face of the card, then 4 must be on the back of the card*" and for this interpretation the best response is to **select Just P** and if the constant anaphor is in **reverse it would be Just Q**. Current literature also indicates that **Not Q choices are likely to be low** unless there exists a falsifying instance or an exception to the rule that does not threaten the rule.[9]

It might be worthwhile to take a little pause to consider what it means to have P and Q on a timeline. Here, rather than study this from a conceptual point of view, one may

[9] If a falsifying instance implies that the rule does not hold as in the rule "if A then 7", then this rule would qualify to be a descriptive rule. A study of the effects of the deontic/descriptive division on the success rate of competent response answers has already been described in section 4.1.1.

look at the implications of having a timeline on any system of events, so the conclusions made are based on general knowledge of sequences in time.

First of all, every living human knows that all events occur in time so every single event has a time tag specifying when it occurred. When two events occur at the same time then they have a simultaneous time tag and this is true when the two events continue for some duration of time. For example, ‘A’ appears on one side of the card and continues for some duration of time that is simultaneous with 4 being on the other side of the card. However, if the two events occur in sequence, then A would appear on one side of the card, and then disappear followed by 4 appearing on the other side of the card. A directional interpretation of this card could imply considering the A on one side, and then going on to consider the 4 on the other side.

A timeline interpretation perfectly explains the low selection of Not Q because it is not an event on the timeline that is followed while reading the rule itself because the events explicitly mentioned are P and Q. Additionally, the strong falsity interpretations require consideration of the work of Jennifer Freyd (1983a; 1983b; 1987; Babcock & Freyd, 1988; Freyd et al., 1988; Freyd and Finke, 1984; Freyd, 1993). Strong falsity is when subjects assume that the Not of the rule is equivalent to the Not of the consequent of the rule. Freyd has shown a phenomenon where subjects are shown one or more still images of different objects at various positions during motion.

Most of these studies showed that subjects “*expect*” the next position to appear and in fact believe it to be the last one they saw. One of these studies conducted by Freyd and Finke (1984) was run by showing subjects three still positions of a rotating rectangle and found that a significant number of subjects chose the position that was one step ahead of the last one they actually saw.

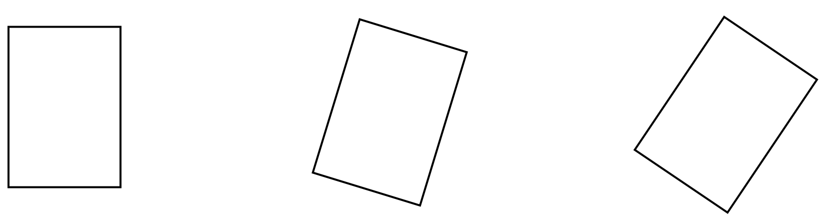


Figure 4.3: A rotating rectangle.

Freyd did not stop there and showed subjects a single image of a man in the air shortly after jumping off a wall, and then showed them two other images that were set temporally apart. One of the two was the same as the one they saw and the other at a slightly different position after the passage of some time (Freyd, 1983b). Results indicated that subjects take longer to identify that the second image is ahead in time which clearly shows that our cognitive systems "*expect*" the next event to occur when we are exposed to a dynamic setting.

Consequently, if we apply the same effect to how subjects consider a rule, then they are expected to consider the event referred to by the antecedent as occurring ahead of that referred to by the consequent. They will "*expect*" the consequent to occur and assume the antecedent to be a given fact that is in the past, so in a sense, this expectation would move their "*present*" point to the midpoint in the rule, cause them to just apply the negation to the second part of the rule. Stenning and van Lambalgen (2004) indicate that some subjects exhibit strong falsity by saying that Not (If P then Q) is equivalent to If P then Not Q.

In this chapter a strict uni-directional temporal sequence will be imposed onto the task in order to identify its effects on reasoning. A new problem format is introduced where subjects are shown a conveyor belt setting that is displayed in figure 4.4 and three positions are shown for the objects mentioned in the rule instead of two. The rule is "*If the striped cube is either on the conveyor belt or in the loading box, then the grey cylinder must be either in the conveyor box or on the conveyor belt.*"

The first issue that requires consideration here is that there are three locations, the striped object can be either in the loading box where items are stored, or on the conveyor belt, while the grey object can be either on the conveyor belt or still in the conveyor box waiting to emerge. Due to the disjunction in the rule, it is important to emphasise that each object is restricted to the above locations causing the linguistic anaphora in the Abstract task to alter its format. Here we do not have the words "*one side*" that can correspond to either face or back, while relating it to the words "*other side*" which would be mapped onto a related side which is the one not occupied as yet. The only location where both objects are allowed to exist is the visible part of the system. Another way of looking at this is as partitioning the location set from "*face*" and "*back*" to "*face*", "*back*" and "*temp*" as shown in figure 4.4. Temp here is the third introduced location that is added to the two classically available in the task.

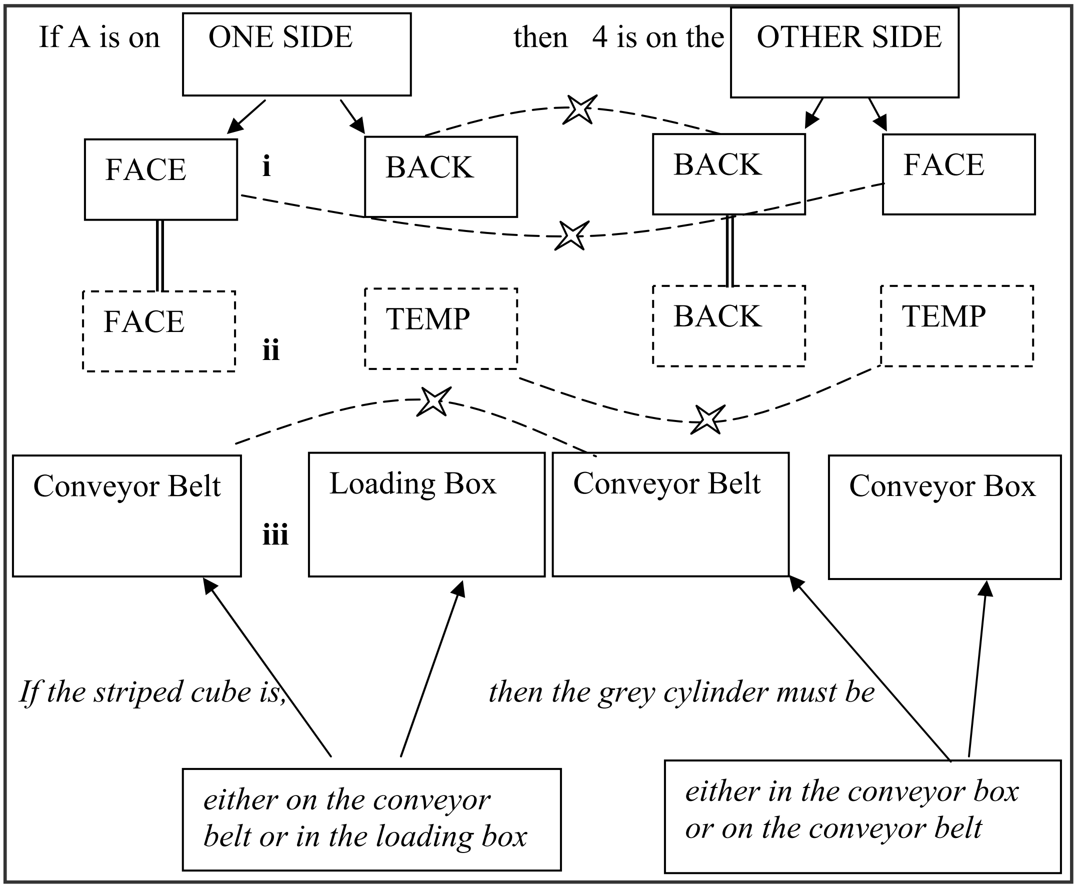



Figure 4.4: A clarification of the anaphora in the Abstract format versus the timeline effect in the Conveyor format

The curved lines shown ensure that when the A occupies the one side slot that corresponds to the face, then 4 cannot possibly also occupy the face of the card. These lines therefore extract the semantic relations that are carried by the anaphora in the "*one side/other side*" wording. By contrast, the line ii shows that once a "*temp*" location is added, then the correspondences are reduced to one restriction instead of two. Additionally, one could described the "*face*" and "*back*" of the cards as "*relative reserved positions*" because once one is occupied by a letter or a number it is reserved and the other remains to describe the location of the number or letter on the other side. By contrast, the conveyor task has an "*absolute reserved position*" because the loading box can only hold a striped object or remain empty, and the conveyor box can only hold a grey object or remain empty. It is this specification of a position reserved for that particular that is expected to result in a higher display of behaviour similar to that of constant anaphora.

# 4.4 Experiments

## 4.4.1 Replication

In order to start off by standing on solid ground, one must start with an experiment that replicates not only the results of selecting P, NotQ but also of the distribution Wason (1968) obtained.

**4.4.1.1 Design Details**

The card task is similar to the classical card problem with A, 11, B and 7 on the cards. The main difference is the use of the word "*must*" which is similar to the card tasks tested by Cheng and Holyoak (1985) and seems to offer no difference in subject choices as the results of replications will show.

**4.4.1.2 Subjects**

44 volunteer students from the University of Bahrain who performed a paper and pen task were given this task ahead of another task. Forty four students performed the Abstract task.

**4.4.1.3 Materials**

The abstract question is as follows: "Below are four cards with letters on one side and numbers on the other side. Your task is to decide which of the cards you need to turn in order to find out whether or not the rule is being followed. The rule is: <u>If a card has an 'A' on one side then it must have a '7' on the other side."</u> Student are then informed to only turn the cards they need to check to be sure. They are shown four cards with A, 11, B and 7.

**4.4.1.4 Results**

The results of this part of the test are shown in table 4.2 below in comparison with Wason's results (1968).

| Abstract | pq | q | p | p¬q | ¬q | ¬pq | pq¬q | ¬p¬q | p¬pq | q¬q | ¬p | p¬p | p¬p¬q | ¬pq¬q | all | none | tot |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Wason 1968** | **10** | 0 | 3 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 16 |
| **Replication** | **20** | 3 | 7 | 2 | 1 | 0 | 2 | 2 | 2 | 2 | 1 | 0 | 0 | 0 | 1 | 1 | 44 |

Table 4.2: A comparison of results obtained by Wason and those obtained at the University of Bahrain

It should be clear from the table that the distribution Wason originally obtained in 1968 is quite replicable, as is evident by the lack of any visible trend or variation from the original Wason results even though these tests were conducted in a completely different country that has a different mother tongue, which is Arabic. The test itself was conducted in English as are all those that follow and the subject base was primarily formed of students who are accustomed to learning almost all their University courses in the English language, but they are by no means as fluent as native speakers. Yet this did not imply that their distribution amongst different possible choices was different from that obtained by Wason (1968). In fact, taking each of the columns above in isolation through a Fisher Exact test shows no significance for any of the categories displayed above without exception.

This is necessary as a starting point from which shifts from one preferred selection to another can be detected as a result of implementing a particular manipulation of the task.

## 4.4.2 Static Conveyor Experiment

Now that a solid foundation has been cast, one can go on to build on it new tasks that have different features than all those in prior work. The aim will be to investigate the effects of modifying the anaphora imposed linguistically by "*one side/other side*" reducing the two constraints imposed into a single constraint. Under this design the card side corresponds to the position occupied by the object and the symbol which is either a letter or a number corresponds to the object and its features which can either be striped or grey.

### 4.4.2.1 Design Details

The conveyor task is somewhat more complex than the abstract task due to the necessary inclusion of a temporal sequence as is dictated by the design targets. This resulted in the design of a conveyor belt scenario that is static at the time of the subjects' examination. The rule is more complex than the abstract because it includes the disjunctive connector 'or'.

The correspondence between the two is shown below.

| | | | | |
|---|---|---|---|---|
| "If | *the striped cube is either on the conveyor belt or in the loading box* | then | *the grey cylinder must be either in the conveyor box or on the conveyor belt* | ." |
| "If | *a card has an A on one side* | then | *it must have a 7 on the other side* | ." |

Figure 4.5: The correspondence between the two rules.

Most students in the Abstract task select P,Q and this behaviour earned a possible explanation offered by Stenning and Van Lambalgen (2004) as described under the "*logic of false*". Subjects can exhibit strong falsity which would cause them to negate *If P then Q* by making it *If P then NotQ*. This would cause them to search for either P or Q as a way of validating the resulting rule, and since this option is not available they may resort to selecting P,Q which is P and Q. The "*expectation*" that can be predicted by Freyd's work (1983b) is that far more selections of Q than those of P are also expected to reduce the effect of strong falsity or the "*logic of false*". In a conveyor system the bias is more towards accepting that the journey will go from P to Q so the probability of having NotQ is reduced. If this happens subjects are less likely to arrive at this interpretation and consequently less likely to make a high percentage of P,Q selections.

Last but not least, we should not overlook the infamous bi-conditional interpretation where subjects interpret the rule "*If P then Q*" as "*If P then Q and If Q then P*". Since we are placing P and Q along a timeline on a conveyor belt this would be interpreted as *If P comes first then Q follows* and *If Q follows then P <u>came first</u>*. From the previous discussion about the directionality of thought, we notice that this statement contains a variety of tense orders that differs from the present-present so common in the task. The difference in the tense of the verb may in fact be a cause to discourage subjects from making a bi-conditional reading of the rule. Unfortunately, this too would result in a lower P,Q selection percentage but this will have to be untangled from the possibility of implying strong falsity within the experiments of this chapter.

This setting therefore tests the effect of strong falsity or a bi-conditional interpretation by suppressing them and predicting a lower selection of P,Q as a result of suppressing them. Simultaneously, it tests the effects of anaphora by partitioning possible locations and making one of them constant. If the results show a bias towards constant anaphora interpretations then indeed this may be the culprit for Just P and Just Q selections.

#### 4.4.2.2 Subjects

41 volunteer students from the University of Bahrain performed a paper and pen task and were randomly given any one of the experiment versions. The number of students who performed the Abstract first format were 21, while 20 performed the Conveyor first format.

#### 4.4.2.3 Materials

The materials were composed of two questions given in alternative order to each of the groups. The abstract question is as follows: “Below are four cards with letters on one side and numbers on the other side. Your task is to decide which of the cards you need to turn in order to find out whether or not the rule is being followed. The rule is: If a card has an ‘A’ on one side then it must have a ‘7’ on the other side.” Students are then informed to only turn the cards they need to check to be sure. They are shown four cards with A, 11, B and 7. This is the exact same question as the one presented above in the replication task. The second question is as follows: “Following are four figures that show four conveyor belt systems. Each transports two objects one at a time. A striped object first exits from within the conveyor box, appears on the belt and then falls into the loading box to be stored. The striped object is followed by a grey object that is either on the conveyor belt or still waiting in the conveyor box.

You are asked to check the following rule by selecting any of the following cases that need inspection. Inspection involves opening the appropriate box and verifying that it contains the appropriate object.

The rule is: If the striped cube is either on the conveyor belt or in the loading box, then the grey cylinder must either be in the conveyor box or on the conveyor belt.” Then they are shown four figures as in Figure 4.6 with the four possible cases each with a checkbox ahead of it and a sentence as in: “Look inside the loading box.”

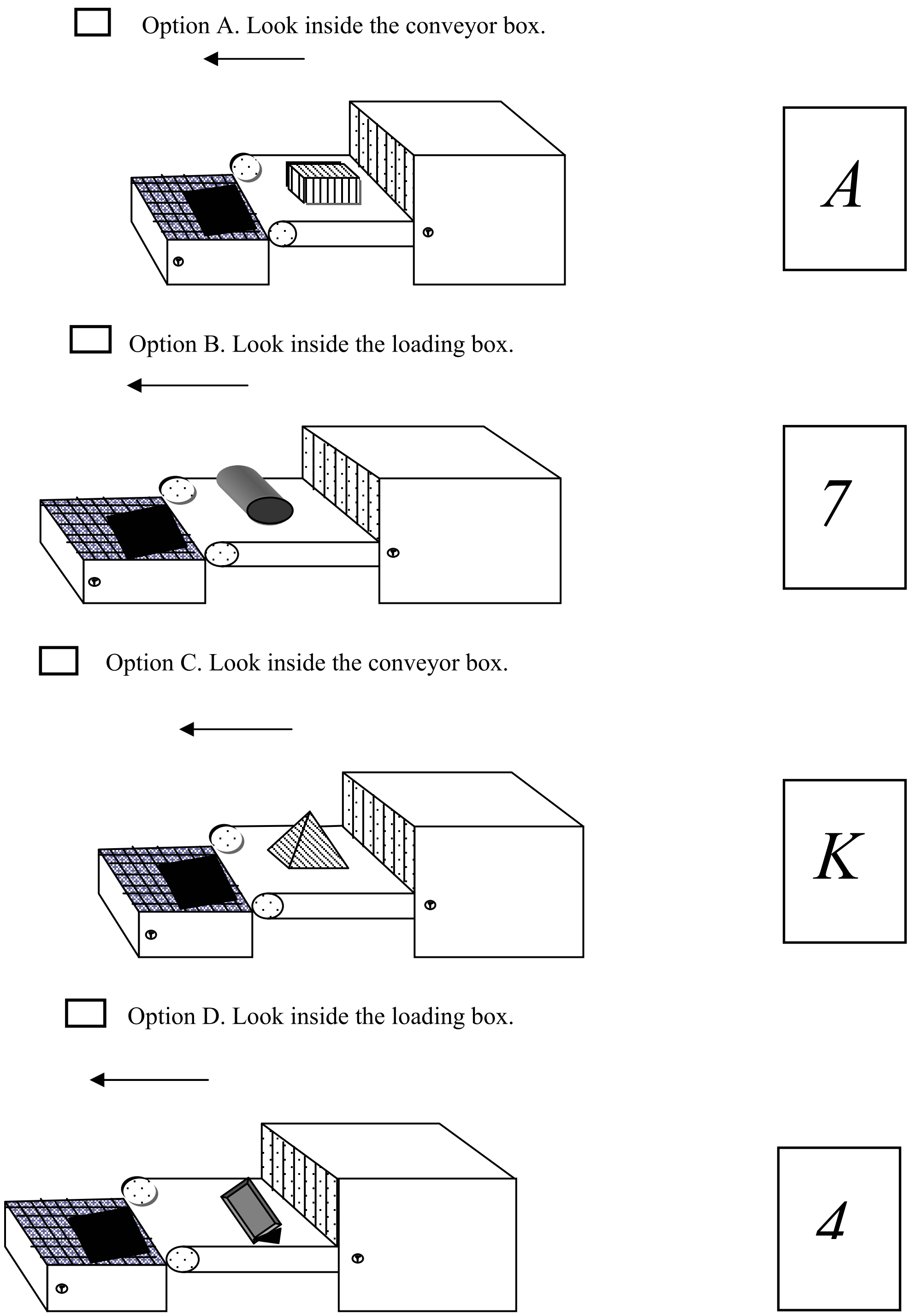



Figure 4.6: Materials for the Static Conveyor Question contrasted against the corresponding cards in the abstract task

#### 4.4.2.4 Results

It is informative to take a look at the card selections to get an idea of the main differences between the two tasks. Additional information is provided in the tables below that show specific card selections when each task is before or after the other task.

| Abstract | pq | q | p | p¬q | ¬q | ¬pq | pq¬q | ¬p¬q | p¬pq | q¬q | ¬p | p¬p | p¬p¬q | ¬pq¬q | all | none | tot |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Abstract first** | **10** | 1 | 2 | 2 | 0 | 0 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 21 |
| **Conveyor first** | **13** | 3 | 3 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 20 |

Table 4.3: Abstract results before and after the static conveyor

| Conveyor | pq | q | p | p¬q | ¬q | ¬pq | pq¬q | ¬p¬q | p¬pq | q¬q | ¬p | p¬p | p¬p¬q | ¬pq¬q | all | none | tot |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Abstract first** | **2** | **4** | **7** | 2 | 1 | 2 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 21 |
| **Conveyor first** | **5** | **7** | **2** | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 20 |

Table 4.4: Static conveyor results before and after the abstract question

The main differences in subject behaviour seems localised in Table 4.4 to the selections of P, Q, Just P and Just Q as per design predictions. However, a Fisher Exact test run on a per category basis shows that only P selections approach significance with a probability less than 0.076.

#### 4.4.2.5 Discussion

There does not seem to exist any effect here due to order and due to the small number of subjects who performed the dual task; this phenomenon is purely localised as when the same task was run on more subjects as was done by Wason (1966, 1968), replication was achieved as was shown in the previous section.

However, the crucial difference between the two tasks is that the conveyor task has an absolute reserved position while the card task has a relative reserved position and the latter dictated by the anaphora. The case of the single restriction implies that conveyor belt position is occupied only by one of the two objects. If a striped cube is on the conveyor belt, and the grey cylinder is in the conveyor box, then the rule is not falsified. Additionally, if the grey cylinder is on the conveyor belt, and the striped cube in the loading box then the rule is not falsified. So these two conditions correspond to the P,Q selection that is so common in the card task. However, what occurs here is that subjects tend to select less P,Q and more of Just Q conditions while keeping the high selections of Just P.

| Conveyor | pq | q | p | p¬q | ¬q | ¬pq | pq¬q | ¬p¬q | p¬pq | q¬q | ¬p | p¬p | p¬p¬q | ¬pq¬q | all | none | tot |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Wason 1968** | **10** | **0** | **3** | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 16 |
| **Static Conveyor** | **5** | **7** | **2** | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 20 |

Table 4.5: Conveyor results contrasted with Wason's results.

The first three columns yet again bid attention because they show a difference in PQ, JustQ but not in JustP selections. If we run a Fisher Exact test on a per column basis we find that the probability of PQ selections occurring as such is less than 0.02, while that of Q is less than 0.009. The largest difference of course is the JustQ selection which is 0 in Wason's (1968) data and 7 which is 35% and in fact larger than both P selections and PQ selections.

A comparison of the data obtained here between the Abstract and the Conveyor tasks is shown in Figure 4.7.

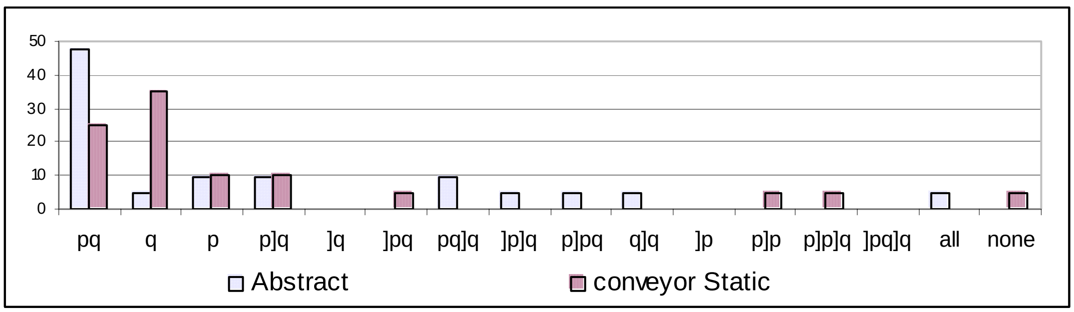



Figure 4.7: Conveyor questions compared to Abstract Question selections (percentages)

The difference in tense when a subject tries to verify the rule in both directions may be one of the causes of a low subject selection of P, Q because it discourages bi-conditional interpretations of the rule. Another characteristic of this format is that it seems to discourage the "*strong falsity*" assumption where subjects would negate a rule by negating only the consequent. The conveyor question is expected to imply an "*expectation*" of the consequent thereby shifting attention more towards the consequent. It is not expected that this type of interpretation should be suppressed as the shift of focus should encourage more of it. The lower P,Q selections therefore, seem to indicate that more than one cause contributes to its popularity amongst student choices, and the suppression of one of the possible causes does not remove its candidacy as a possible choice.

In order to also evaluate how the analogy between the constant anaphoric reading in the abstract task and the Just P and Just Q selections in the conveyor task works, one may look at the correspondence.

Just P selections:

If the card has an A on its face, then it must have a 7 on the back.

If the striped cube is on the conveyor belt, then the grey cylinder must be in the conveyor box.

Just Q selections:

If the card has a 7 on its face, then it must have an A on the back.

If the grey cylinder is on the conveyor belt, then the striped cube must be in the loading box.

Notice that the position corresponding to the "*face*" position can hold the two possible options, while the one corresponding to "*back*" is well distinguished in the first as compared to the second possibilities. In a sense, the distinguishing factor between the two options is what is on the conveyor belt which happens to be always visible. Consequently, the only position that may contain one of the two objects is visible and constant, while the others are reserved for whatever is not accommodated onto the visible position. If the visibility of a position that may hold either of two objects is constant, in the four options shown, then it may induce an interpretation of the Just P or Just Q selections shown above, because it will be the focus of subject attention, and other choices revolve around it.

Additionally, if we add to this the likelihood of "*expecting*" the next event to occur in a conveyor setting, then subjects are expected to make more Just Q selections than they would make Just P selections as is the case here.

Notice that the Conveyor question is more susceptible to sequence effects than the Abstract question because while the Abstract did not allow the Conveyor question to make any serious alterations in behaviour, the Conveyor question did exhibit an almost significant difference when the Abstract question comes first. So what makes this question form "*weaker*" (if one may use this word) when it comes to sequence effects? The main difference between the two tasks is that the two objects arrive at the central belt at different locations in time. A possible answer to this is that the Conveyor question is displayed graphically so perhaps that graphical depiction of an object on a system capable of moving may in fact prompt subjects to expect a change or alteration to occur. If this change is expected along the direction of reasoning, then it may cause a readiness to alter selections that does not exist when the setting does not imply any motion as in the case of the Abstract task. This is similar when reasoning arrives at a conclusion that is considered satisfactory versus when reasoning is still in progress and sensitive to external effects. Possibly implying motion makes the reasoning process more sensitive towards effects because it is under the illusion of change.

However, one cannot be satisfied with these results without further tests for the "*dynamic*" nature of the conveyor question and the disjunction that was necessary for this task.

## 4.4.3 Implied Dynamics Conveyor Experiment

The experiment displayed above differs from this one in two main ways: this one explicitly informs subjects that the conveyor is in motion while the other implies it is static and that this one has no disjunction to restrict locations.

### 4.4.3.1 Design Details

This experiment is an accurate copy of the card task given in the first experiment. The difference here is that motion will be implied verbally both through the rule as well the option labels on the four selections. If we compare this rule with the prior version, the following results.

| | | | | |
|---|---|---|---|---|
| "If | *the striped cube is* **either on the conveyor belt or in the loading box** | then | *the grey cylinder must be* **either in the conveyor box or on the conveyor belt** | ." |
| "If | *the striped cube is* **in the conveyor belt system** | then | *the grey cylinder must be* **the grey object directly following it as the conveyor belt moves**. | ." |

Figure 4.8: The correspondence between the two rules

There are several main differences between this format and the one used in the prior experiment. The first and clearest is that the disjunction disappears, while still implying the same "*absolute reserved position criteria*" found in the previous section. Another issue is the change that occurs to the consequent as it is linked to the antecedent through an extra description that shows a "*direct follow*" relationship between the two objects. Additionally, the choices that are offered include two choices that include the words: "*wait for the next object to appear*" and two choices that are similar to the previous task that state: "*look into the loading box*".

If the cause of the selections of Just P and Just Q made in the static conveyor is imposed due to the "*reserved absolute position*" that is also clearly shown here, because of the dynamic scenario they should still be as high as the ones made in the case of the static conveyor. Relating the second object to the first through an added constraint should increase subject focus on the consequent, especially since the constraint shows a sequencing that it is natural to "*expect*" given a dynamic conveyor scenario. By contrast, more subject doubt should result because the possible locations for each object are not explicit in the rule.

#### 4.4.3.2 Subjects

45 volunteer students from the University of Bahrain performed a paper and pen task and were randomly given any one of the experiment versions. The number of students who performed the Abstract first format was 23, while 22 performed the Conveyor first format.

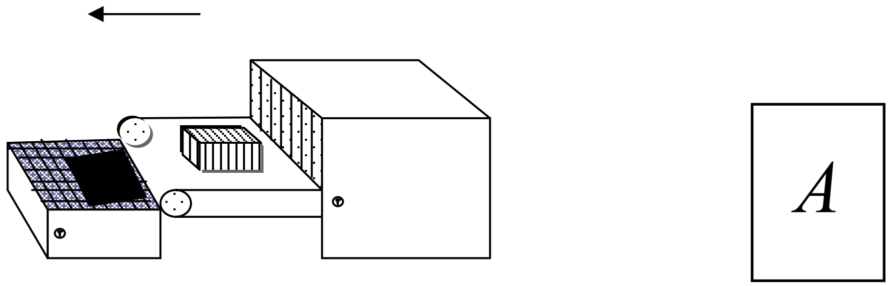
#### 4.4.3.3 Materials

The materials for this task are composed of two questions given in alternative order to each of the groups. The abstract question is the same as the one given in section 4.3.2.3. The second question followed the design described in the previous section and has the exact wording of the question given in 4.3.2.3 with the exception of the rule, that is as follows:

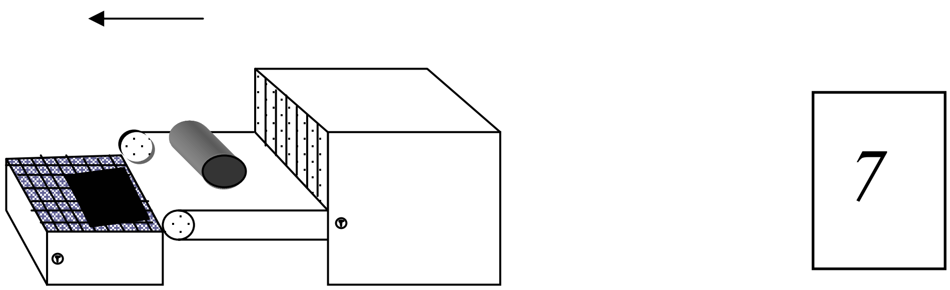
The rule is: <u>If the striped cube is in the conveyor belt system, then the grey cylinder must be the grey object directly following it as the conveyor belt moves.</u>

They are then given the same diagrams of the conveyor system but the sentence to check the storage box in two of the cases is replaced with: "*Wait for the next object to appear.*"

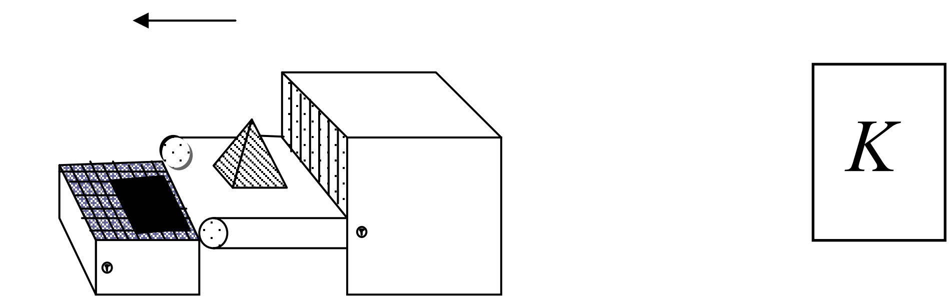
☐ Option A. Wait for the next object to appear.




☐ Option B. Look inside the loading box.




☐ Option C. Wait for the next object to appear.




☐ Option D. Look inside the loading box.

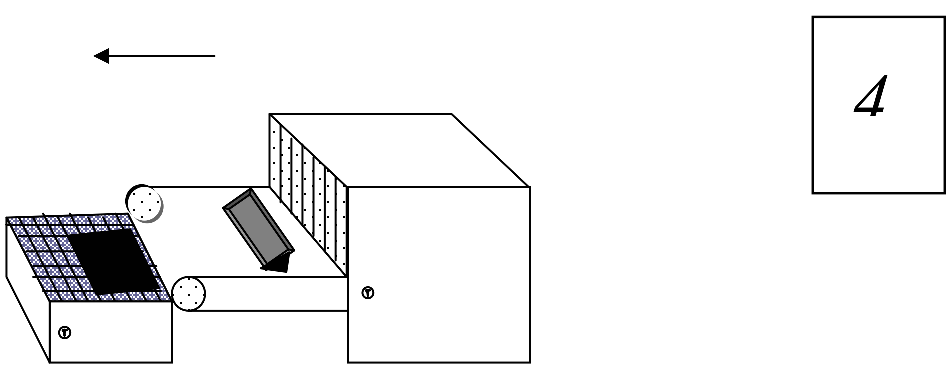



Figure 4.9: Materials for the Dynamic Conveyor Question

#### 4.4.3.4 Results

Results are shown in tables 4.6 and 4.7.

| **Abstract** | **pq** | **q** | **p** | **p¬q** | **¬q** | **¬pq** | **pq¬q** | **¬p¬q** | **p¬pq** | **q¬q** | **¬p** | **p¬p** | **p¬p¬q** | **¬pq¬q** | **all** | **none** | **tot** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Abstract first** | 10 | 2 | 5 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 23 |
| **conveyor first** | 10 | 0 | 6 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 1 | 22 |

Table 4.6: Abstract results before and after the static conveyor

| **Conveyor** | **pq** | **q** | **p** | **p¬q** | **¬q** | **¬pq** | **pq¬q** | **¬p¬q** | **p¬pq** | **q¬q** | **¬p** | **p¬p** | **p¬p¬q** | **¬pq¬q** | **all** | **none** | **tot** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Abstract first** | 1 | 8 | 5 | 1 | 0 | 0 | 0 | 0 | 1 | 5 | 1 | 1 | 0 | 0 | 0 | 0 | 23 |
| **conveyor first** | **3** | **5** | **4** | 2 | 0 | 1 | 0 | 0 | 0 | 5 | 0 | 1 | 0 | 0 | 1 | 0 | 22 |

Table 4.7: Conveyor Question selections in both conditions

The Implied Dynamic Conveyor results shown here mimic those of the Static Conveyor in their distribution between Just P and Just Q choices with a low selection rate of PQ in spite of the fact that this rule has no disjunction. But for added clarity we can compare the results directly between the static conveyor first, implied dynamic conveyor first and Wason (1968) results in Table 4.8.

| **Conveyor** | **pq** | **q** | **p** | **p¬q** | **¬q** | **¬pq** | **pq¬q** | **¬p¬q** | **p¬pq** | **q¬q** | **¬p** | **p¬p** | **p¬p¬q** | **¬pq¬q** | **all** | **none** | **tot** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Static Conveyor** | **5** | **7** | **2** | 2 | 0 | 1 | 0 | 0 | 0 | **0** | 0 | 1 | 1 | 0 | 0 | 1 | 20 |
| **Implied Dynamic Conveyor** | **3** | **5** | **4** | 2 | 0 | 1 | 0 | 0 | 0 | **5** | 0 | 1 | 0 | 0 | 1 | 0 | 22 |
| **Wason (1968)** | **10** | **0** | **3** | 1 | 0 | 0 | 1 | 0 | 0 | **0** | 0 | 1 | 0 | 0 | 0 | 0 | 16 |

Table 4.8: Conveyor Question selections of both questions compared to Wason's results (1968)

A Fisher Exact test shows no significant differences between the two conveyor formats except in the Q, Not Q selections. While it shows significance in the PQ and Q, and Q,Not Q selections from Wason (1968) with a probability of 0.053, 0.052 and 0.052.

#### 4.4.3.5 Discussion

This task shows no serious differences in results from the one presented in the previous section as predicted even though it does not have a disjunction and relies on the imposed explicit sequence in clarifying the possible positions of objects. The main difference, therefore is the added description that is attached to the consequent of the rule which is expected to shift subject focus onto the consequent of the rule.

Therefore, it is not at all surprising that more subjects select Q, Not Q selections in this format than in the previous one. In a sense, they are selecting all possible choices, after "*expecting*" the conveyor to have moved to the consequent time. This behaviour is similar to those who adopt the "*logic of true*" for the abstract card task, except here the scenario shifts their focus towards the consequent so the only two options that require testing are Q and Not Q.

In spite of the clear results shown so far, one cannot help wondering about the true effects of the disjunction in the static format of the conveyor experiment.

### 4.4.4 Ambiguities and the issue of "*or*"

It may be clear that logical interpretations are fortified or weakened by experimental formats of the conveyor task. This does not imply, however, that a shadow does not linger on making one wonder about the effects of removing the disjunctive "*or*" from the static conveyor setting.

#### 4.4.4.1 Design Details

It would be nice if one could add a level of ambiguity to the static conveyor such that the effects captured in section 4.3.2 are reversed and those categories returned to Wason's (1968) range of values. Only through this form of reversal can one make solid conclusions of the effects captured in the static conveyor experiment.

Since the effects were elicited through specifying "*absolute reserved positions*" versus the "*relative reserved positions*" as implied by the variable anaphora that exist in the abstract form of the task, doubt must be cast on these positions. The best way to cast doubt is to remove the disjunction that clearly specifies the possibilities; after all, the variable anaphora imposes two constraints rather than the one imposed in this problem and leaves two positions ambiguous. When the disjunction is removed, subjects may consider it possible here for each of the two objects to occupy any of the three locations available.

#### 4.4.4.2 Subjects

31 volunteer students from the University of Bahrain performed a paper and pen task.

#### 4.4.4.3 Materials

The materials were composed of the conveyor question given in the following format: "Following are four figures that show four conveyor belt systems. Each transports two objects, one at a time. A striped object first exits from within the conveyor box, appears on the belt and then falls into the loading box to be stored. The striped object is followed by a grey object that is either on the conveyor belt or still waiting in the conveyor box.

You are asked to check the following rule by selecting any of the following cases that need inspection. Inspection involves opening the appropriate box and verifying that it contains the appropriate object.

The rule is: <u>If the striped cube is in the conveyor belt system, then the grey cylinder must be the grey object also in the system.</u> Then they are shown four figures with the four possible cases each with a checkbox ahead of it and a sentence such as: "Look inside the loading box."

Students were then shown what is depicted in Figure 4.6. The corresponding card task selections are shown next to each conveyor.

#### 4.4.4.4 Results

A clearer analysis of subject selections is shown in Table 4.10.

| **Conveyor** | **pq** | **q** | **p** | **p¬q** | **¬q** | **¬pq** | **pq¬q** | **¬p¬q** | **p¬pq** | **q¬q** | **¬p** | **p¬p** | **p¬p¬q** | **¬pq¬q** | all | **none** | **tot** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Wason 1968** | **10** | **0** | **3** | 1 | 0 | 0 | 1 | 0 | **0** | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 16 |
| **Static Conveyor** | 5 | 7 | 2 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 20 |
| **Without "or"** | **11** | **4** | **1** | 0 | 1 | 0 | 0 | 1 | **7** | 0 | 2 | 0 | 0 | 0 | 4 | 0 | 31 |

Table 4.9: Wason's (1968) results and both Conveyor Questions

First of all a comparison between the two static conveyor questions reveals that they do not differ from each other significantly when using the Fisher Exact test on a per category basis except with choices of P, Not P, Q and it appears that these are more likely when the level of ambiguity is high enough to introduce the possibility of brittleness and exceptions to the rule. This is expected because the lack of the disjunction makes it possible for subjects to erroneously assume that the striped object could be in either of the three locations and the same for the grey object, thereby losing the restriction explicitly imposed on the static version of the task.

Additionally, there are no significant differences between this version of the task and Wason's (1968) using a Fisher Exact test in any of the categories shown except with respect to choices of P, NotP, Q.

For the sake of thoroughness we look at card choices.

| | P | **Not P** | Q | Not Q |
|---|---|---|---|---|
| Static Conveyor Question | 11(55%) | **3(15%)** | 13(65%) | 3(15%) |
| Static Question without OR | 23(74%) | **14(45%)** | 26(84%) | 6(19.3%) |

Table 4.10: Abstract Question selections in both conditions

The overall distribution shown in table 4.11 shows some variations in subject behaviour between the task version that contains the disjunction versus the one that does not contain the disjunction. A Fisher's Exact test to measure the probability of selecting Not P shows a significance of 0.019 in difference between the two tasks. As predicted the percentage of selecting Not P is significantly higher when the disjunction is not part of the rule.

#### 4.4.4.5 Discussion

Sure enough, once the "*absolute reserved position*" is not clearly reserved for the appropriate object, and the striped object could appear in any of the three locations and the same goes for the grey object, performance in the first three categories goes back to conform with the abstract form.

The other difference which is a high selection of P, NotP, Q reflects that this rule differs from the abstract in the number of possible locations, namely three instead of two. This may increase subjects' perception of the brittleness of the rule which is the cause Stenning and van Lambalgen (2004) suggested for such a choice.

The lack of clarity of the positions is further exhibited by the high selection rate of the Not P card at 45% when compared to the static conveyor task 15%, or the implied dynamic conveyor task at 4.54%. This shows a clear association between the probability of selecting this card and the level of the ambiguity exhibited by the task.

## 4.5 Discussion

There are two main issues that require attention in the experiments presented in this chapter that stem from the results obtained. The first is a comparison between the two tasks and the second is an interpretation of the findings.

A comparison between the two tasks yields a number of differences that may or may not have direct implications on subject card choices. One of these is the use of the indefinite article "*a*" in the abstract/descriptive task while using the definite article "*the*" in the conveyor setting. Manktelow and Evans (1979) tested the rule "*If the letter is N, then the number is 3*" and if their results are compared to Wason's (1968) distribution no significant difference results in subject distributions using a Fisher Exact test on a column by column basis. Evidently, on its own this difference does not play a major role in influencing results, however, from a semantic point of view the word "*the*" limits the world of possibilities to "*the cube described in the rule*" rather than any other. By contrast "*an A*" only describes an instance of the letter A without specifying which particular instance that is, so any A on that card would do even if it is different from the one specified in the rule.

The use of the definite article, therefore, may possibly be interpreted by subjects to imply the visible striped cube. This, in a sense, is comparable to the rule "*If you turn the A, then you will find a 4*". If this rule is given to subjects along with the cards A, K, 4, 7 then it may be interpreted to imply that turning the card with "*A*" on the face is sufficient to check the rule. In this case, subjects would interpret the words "*the A*" as describing "*the visible A*" on the face of the card.

However, one should recall that the context of the conveyor setting describes a relationship between a striped object and a grey object using indefinite articles. Consequently, when the definite article is used in the rule, the aim is to reinforce that this striped cube is a member of the set restricted by that relationship and sure enough P selections in the conveyor task are no higher than those made in the abstract/descriptive task.

Another issue of comparison is that in the static conveyor question the possible positions of the objects are explicitly stated while in others implicitly restricted. The loading box, for example, can only hold either the striped cube or remain empty which qualifies it to be described as an "*absolute reserved position*". Another perspective on the same setting or comparison can be considered from a semantic point of view.

In the abstract task, the reference of "*one side*" affects the reference of "*other side*", which qualifies the restriction to be anaphoric. The "*one side*" quantifier has a domain set that may include either the set (A, K) or the complement sets (7, 11). This in turn affects what the quantifier "*other side*" refers to, such that it would represent the set other than the one referred to by "*one side*".

By contrast, the conveyor question has different referential features. A strongly imposed restriction is the one represented in the sequence of having one object precede the other temporally in a conveyor setting. Consequently, it imposes a unidirectional relationship such that the relationship is always striped followed by grey. Semantic sequence is represented in constructs that assume presuppositions, as in the following example:

We suppose that *Max has children*

"*If Max has childen, Max's children are American."* (Fauconnier, 1994).

S P

"*Max's children" in this context is an explicit presupposition and* S has to be established in the domain M before P is established in it. The conveyor task describes four conveyor systems, where a striped object is always followed by a grey object. Another way of looking at this is that if there is a grey object, then there is no doubt that it was preceded by a striped object, as we only doubt what the striped object was. In other words, the existence of a grey object on the belt carries a "*presupposition*" that a striped object already passed.

The consequent, "*Max's children are American*" presupposes that Max has children, whether the precedent is read or not. If we apply the same reading to the conveyor question, we find that "*the visible grey pyramid*" will presuppose that a striped object passed ahead of it and this should result in high Just Q selections as is the case in these experiments.

To sum up, the experiments presented in this chapter attempted to replace the anaphoric "*one side/other side*" where the referential link of one restricts the referential link of the second, with a unidirectional restriction, where one object always precedes the other. So the comparison is between two referential links that each affect the other in a specific way, to a unidirectional link where one reference is always ahead of the other. If subject behaviour is affected by that anaphoric link, then results should reflect the difference in the semantic implications of the conveyor task. Sure enough, results show a high selection of Just Q which is predicted through "*presupposing*" that the P object has already passed through the conveyor if Q is encountered.

This chapter, therefore, studied the semantic restrictions imposed within the anaphora in the abstract/descriptive question through altering it and comparing results to what is predicted through semantics. The task links values obtained from a set of letters (A, K) to a set of numbers (7,11) and now that the relation established between them has been explored, it would be wise to consider the effects of placing the set partition between the antecedent and consequent. If two colours are added, then two emerging groups of those that are white and those that are black, result. Partitioning, then becomes possible for the two above sets in various ways in order to identify the effects of the number of sets as well as partitioning on subject behaviour. This will be done in chapter 5.

# Chapter 5

# Contrasting Sets, Complement Sets and Logical Negations

The study of a problem as large and complex as this, can only be achieved through partitioning and sub-partitioning of the problem's components. The existing difference that has been discussed in Chapter 3 is solely between deontic materials that have achieved high percentages of the competent response and descriptive materials that result in low percentages of competent response choices. Stenning and van Lambalgen (2004) recently offered insight on some of the interpretations subjects may be assuming when making these choices. Chapter 4 studies the relationship imposed by the pronouns of anaphora "*one side/other side*" onto what they reference from the set of letters and complement set of numbers. This is done through an alternative quantifier/referential relationship imposing a time sequence where the member of the set is always assigned directly ahead of the member of the complement set. This chapter by contrast, leaves the quantifier/reference restrictions constant while studying the relationship between the set of letters and its complement set of numbers.

In the abstract task, one is told that the letter set once referenced by the quantifier "*one side*" would lead the anaphor "*other side*" to reference the number set. This is the logical negative of supposing a member of the set of letters references the other member of the same set which is composed here of only two members. For example, if S is a proposition that there is an A is on one side then the expected Not S is that there is a K on one side.

However, to date, no study has investigated whether subjects' interpretations of this logical negative is as is predicted by researchers. These interpretations may in turn be affected by how subjects react to different foreground/ background rule combinations through the use of colour. Colour can be used to emphasise various distinctions without altering propositional content. It is argued that contribution to meaning occurs through several pragmatic factors including implicature, presupposition and as is the case here "*Information Packaging*". This is described as a structuring of sentences by syntactic, prosodic or morphological means that arises from the need to meet communicative demands of a particular context or discourse. In particular, information packaging indicates how information conveyed by linguistic means fits into the (hearer's mental model of the) context or discourse. (Vallduvi and Engdahl, 1996)

Here the addition is implemented by partitioning the set of letters and the set of numbers into what is white and what is black. Various combinations can then be constructed so that both the letter and number are of the same colour versus having the set and complement set division emphasised through colour. This has the effect of increasing the number of possibilities from 4 to 8 but this is not a central concern as will be shown in the discussion section of this chapter.

The added dimension of variation alters the way letters and numbers are grouped by students who respond to the question through making their colour feature dictate which group they belong to rather than relying on the differences that exist between letters and numbers. Consequently, the semantic attributes here are emphasised and utilized to group the set members while these attributes exist for members of both sets of letters and numbers.

Additional tests are then attempted in order to identify if colour is capable of replacing the other set of numbers in a way that would not influence subject behaviour. The aim here is to alter the relationship between what is referred to by the antecedent from what is referred to by the consequent.

One point worth noting at this stage is that the common sets investigated by Wason (1968) in this task are primarily letters versus numbers, vowels versus consonants and even versus odd. These sets have a unique hierarchical structure, which is not the case when the colours black and white are added. Colour is a feature that both letters and numbers can have and therefore allows multiple hierarchies. For example, odd letters and even numbers can be both white, placing both into the set of white values. Another example is when letters can appear in both white and black and numbers can appear in both white and black while another example is when both letters appear in white while both numbers appear in black. The latter colour distribution reinforces the distinction between a letter and a number.

To sum up, this chapter will explore the effects of emphasising various contrasts between the various sets such as the letter set and number set through partitioning them through the colours white and black, i.e. information packaging. As a control, it will explore cases where colour is part of the truth conditional semantics. Last but not least, the chapter will study subjects' interpretations of the logical negation of the antecedent and the consequent.

## 5.1 Colour Partitioning and Information Packaging

The abstract form of the task will be examined through various means; the first is to investigate the assumption made earlier in this work that subjects do indeed impose different logical interpretations onto the task as is illustrated by Stenning and van Lambalgen (2004). Fillenbaum (1978) observed that in 30% of conditionals similar to the selection task rule, subjects would give "*If P then Not Q*" as the negative of the rule "*If P then Q*". Stenning and van Lambalgen (2004) showed that this would lead subjects to select "*either P or Q*" as the optimal answer to the selection task question. Since this is not a possible answer, they may resort to selecting the P, Q cards.

The work presented here will further investigate this by asking subjects to report what they regard as the negation of the antecedent and the consequent in isolation, with the goal of eliciting different responses from different subjects and investigating whether interpretations correspond in any way to conclusions drawn.

However, studying how subjects view negatives is hardly sufficient to exhaust the semantic relationships between the antecedent and consequent reference, versus the semantic relationships between the antecedent and its complement for example. In order to study contrasts between them, it must be possible to alter the groups that the letters and numbers belong to such as to ensure they are distinct from each other. Since there are two sets, one of letters and the other of numbers, that are represented in a very specific fashion, it is difficult to group a member of the number set with a member of a number set without altering their information describing them. One way around this problem is by adding a distinguishing feature that would redefine these as members of two distinct groups. The main characteristic required of such a feature is that it can describe the letters and numbers shown on the cards, while carrying a trait of contrast to test these relationships within the task.

The contrast between white and black offers itself as a tool that can be used to partition the abstract task materials. For example, the letter set used in this thesis is composed of the letters A and K. Here they can either both be white, both black or one white and the other black as per the requirements of the test. In a similar fashion, both the antecedent and consequent can refer to things that are white while their logical negatives are left as black, or the antecedent's referent can be white while the consequent's is black. Due to the large number of emerging possibilities only some tests were run as part of this thesis while leaving other possible combinations that the reader can easily formulate for future work.

### 5.1.1 Different Interpretations of Logical Negatives

As has been previously mentioned, the experiments here contain an additional component that probes what subjects regard as the negation of the antecedent and consequent. Each task has a page that follows the main page, containing a question that is only attempted after solving the selection task question. Students were asked to complete the task in pen such that any alteration to the first page which contains the abstract task would be detected and the results would be disregarded. They are asked to fill in the values of "*the negation*" of each of the antecedent and consequent. For example, if the rule is of the form *If P then Q,* then the question would ask subjects; If S = "*A is on one side*" Not S = ______ and expect them to write what value they associated with their interpretation of the "*logical not*" of the premise. This gives us some evidence about how subjects interpret negations in this task. Of course, it also introduces issues about how they interpret this new task, to which we will return.

Perhaps at this point in time a question may arise as to the rationale behind such a test. It has been to date taken for granted that the negative of the case when the antecedent referring to the value A appearing in one side, is when a case arises that K is to appear. Here, the K is the member of the letter set that does not appear in the rule and is obtained through isolating the antecedent clause from the conditional structure and regarding the negation independently. However, given current research in the field including the work of Stenning and van Lambalgen (2004) things are not always the way they seem. Subjects were shown to impose various interpretations onto the selection task that affects their reasoning. If this is the case when reasoning about the task, then it may also be the case that in what may be a simple negation task, subjects impose preconceptions that may influence the conclusions they arrive at. Some of the logical preconceptions include the assumption that the negation task is not independent of the rule causing subjects to relate the two. Since this possibility has not been considered to date it seemed worthwhile investigating it during this work, as it may illuminate possible causes behind the ambiguity in the task.

### 5.1.2 Designing the Exploration Path of the Descriptive Task

In designing the following experiments, one should be able to select some of the different partitions from the large number of possibilities marked by colour according to some criteria. First of all, there must be one condition used to replicate, and some to test the descriptive letter/number task.

Starting from Wason's (1968) original task as a baseline, colour can be added in two ways: either by making two of AK47 letters and numbers always black and the other two always white, or to use just two letters A and B which can be black or white without any numbers. The first way is through partitioning the sets and complement sets while the second is through reducing the number of sets by replacing the set of numbers with that of colours. One may note a difference between the two in that the first approach, while it displays the four cards to subjects, is unable to exhibit the full range of possibilities to them on just four cards, while the second displays the full range of possibilities as is the case on the original task. The basic difference is whether subjects make a restrictive or attributive reading of the clauses. For example, if a clause is of the form "*This is an A which is white*" then the clause "*which is white*" may restrict the number of As that are described in the clause to those that are white while there also exist in the same world, many As that are not white. On the other hand, the same clause may also be read that the colour white is an attribute or trait that is used to describe the letter A. The assumed reading here is the second and results of the negation task will support this assumption by showing that an insignificant number of subjects exhibit this restrictive reading.

Other issues that come up during the comparison will be discussed in the appropriate section below.

#### 5.1.2.1 Replication

The first condition of course is a baseline task that aims at replicating Wason's (1968) original results.

**Condition 1:** If there is an **A** on one side of the card, then there is a **7** on the other side of the card.

**Background rule:** On each card, there is a number on one of its sides and a letter on the other.

An association between the interpretations of logical negatives with subject card choices may emerge even if data are too sparse to show this in all cases. In the replication task, a comparison of the frequencies of choices in each category is compared in order to identify if subject distribution is replicable and consequently stable. If this is the case, then one can go on to test other conditions on the same population in order to identify conditions that affect various choices. Wason assumes a classical interpretation of negation, so this chapter will look at other logical interpretations that subjects may impose onto the task in order to perform this operation.

#### 5.1.2.2 Colour Partitioning of the AK47 Rule

In addition to replication, three conditions of the AK47 rule, with different assignments of colour were constructed. In condition 2, colours impose a division between the true values of the clauses of the rule and other values by making both A and 7 which appear in the rule white and leaving other possible choices as black. This keeps the problem the same with respect to the truth conditions when compared to the original Wason (1968) version. The difference it has with the baseline is that it imposes a form of information packaging onto the data such that one type of letter and one type of number is grouped by a colour feature while the other type of each is part of another group in the background rule. The baseline condition has only the letters as a set while the numbers form a complement set.

Condition 3 uses colour to partition the antecedent content versus consequent content such that letters are always black and numbers always white. In this sense colour also marks the "*face*" or "*back*" of the card.

We can then compare condition 2 to condition 3 above to find that they are different such that in condition 2 colour partitions each of the set and complement sets into what is white and what is black, while in condition 3 the colour partitioning is between the set and complement set. Additionally, in condition 2 colour aligns with truth values of the conditional while condition 3 has the antecedent refer to a value of one colour while the consequent refers to a value of another colour. Condition 3 also has another feature which is that colour marks the "*face*" or "*back*" of the card in addition to it being a letter or number.

Condition 4 can then be compared to conditions 2 and 3. First of all, it is similar to condition 2 in that the background rule partitions each of the letter and number sets such that part of each is white while the other part is black. They are different in that condition 2 has the colours aligned with the truth values of the conditional A and 7 while condition 4 is misaligned with them.

Condition 4 is similar to condition 3 in that in both cases the foreground rule has both colours. It differs from it in that the order of the colours is different. Another difference is that in condition 4 the two sets are partitioned into what is black and what is white while in condition 3 the distinction is between set and complement set. Additionally, in condition 3 colour marks what can be on the "*face*" of the card versus what can be on the "*back*", in addition to whether they are from the set of letters or from the set of numbers.

As mentioned above, the cards exhibited for conditions 2, 3 and 4 are only 4 out of 8 possible cards. Table 5.1 shows the main details of the described conditions.

| **Condition** | **Foreground Rule** | **Cards** | **Background Rule** | **Colour Distinction** |
|---|---|---|---|---|
| C. 2 | White A→ White 7 | white A, white 7, black K, black 4 | white(odd numbers + vowels) | Truth/falsity (this implies that both set and complement set are partitioned through colour) |
| C. 3 | Black A→ White 7 | black A, white 7, black K, white 4 | white (numbers) / black (letters) | Antecedent/Consequent (this implies that the set is distinct in colour from the complement set) |
| C. 4 | White A→ Black 7 | white A, black 7, black K, white 4 | white (even numbers + vowels) | Colour marks neither truth/ falsity nor antecedent/ consequent (this implies that both set and complement set are partitioned through colour) |

Table 5.1: The essential details of the three variations of the task where the type of symbol (letter/numbers) is replaced by colour

The complete background rules for the conditions are as follows:

> C. 2: *On each card, there is a number on one of its sides and a letter on the other. Notice that odd numbers and vowels are white, while even numbers and consonants are black.*
>
> C. 3: *On each card, there is a number on one of its sides and a letter on the other. Notice that numbers are white, while letters are black.*
>
> C. 4: *On each card, there is a number on one of its sides and a letter on the other. Notice that even numbers and vowels are white, while odd numbers and consonants are black.*

The reader should notice that the conditions described above all impose an "*information packaging*" grouping to the rules such that it does not alter what causes the rules to be true or false. However, this procedure emphasises various distinctions that may or may not be imposed by subjects. Condition 2 tests if subjects impose a distinction on the cases that fit the rule versus those that do not fit the rule, so if the distinction is already present, results would show no significant differences in behaviour. Condition 3 emphasises the distinction between the values that lie on both sides of the cards, which will show the effects of distinguishing the referents of the antecedent and consequent through colour.

These conditions will test various "*information packaging*" possibilities by adding colour and this in itself raises a question of what would happen if colour replaced one of the sets.

**5.1.2.3 What Happens When Colour Replaces the Number/Letter Distinction?**

The three conditions described above are similar to the baseline in that the conditional relates a member of the letter set to a member of the number set. It is different because colour is added as a feature that members of either of the two sets can have. In the three conditions here, the set of numbers is eliminated to make the conditional relate two members of the set of letters, namely A and B in one case and a member to itself in the other. In other words, colour (Black/White) replaces the letter/number distinction made to the face versus back of the card. This implies that the letters referred to by the antecedent and consequent could either be different or identical. It also implies that in these cases colour does make a difference in the truth versus falsity of the rule.

Condition 6 is similar to condition 3 in the previous section with the exception of utilising one set of letters from which the letters A and B are taken. The background rule is identical in both cases. In condition 6 the full range of possible cases is shown to subjects while in condition 3 only 4 possibilities out of 8 are included.

Condition 5, on the other hand, utilises only one letter in the rule, such that it aligns the letter identity with truth so A appears in both the antecedent and consequent. Therefore, the letter B in both its colour formats is linked to falsity. Condition 7 is comparable to condition 6 with one difference, which is the use of the colour red to replace black and the colour orange to replace white. Red and orange are both easily classified as shades from the group of "*reds*" while black and white are easily accepted as opposites. In conditions 6 and 7, colour does affect the truth/falsity of the rule;

consequently, it will test contrasting the value referred to by the antecedent versus that referred to by the consequent versus referring to two values that have a common grouping as that here of both being shades of the "*red*" colours.

| Condition | Foreground Rule | Cards | Background Rule | Information Packaging |
|---|---|---|---|---|
| C. 5 | Black A→ White A | black A, white B, black B, white A | Black (letters) / white (letters) | Same letter that appears in the antecedent appears in the consequent. |
| C. 6 | Black A→ White B | black A, white B, black B, white A | Black (letters) / white (letters) | The letter in the antecedent is different from that in the consequent. |
| C. 7 | Red A→ Orange B | red A, orange B, red B, orange A | Red (letters) / orange (letters) | The letter in the antecedent is different from that in the consequent. |

Table 5.2: This table shows the essential details of the three variations of the AK47 task

Table 5.2 shows that the main difference between the previous three conditions and conditions 5, 6 and 7 is that the latter three reduce the number of sets into one where a member is referenced by the antecedent and another by the consequent.

These three conditions alter the hierarchical relationship exhibited in the first four conditions. The antecedent in the first four refers to a letter while the consequent refers to a number which is a member of a distinct set while in the case of these three both the antecedent and consequent refer to members of the letter set. Condition 5 tests if the rule tests two different attributes of the same letter, while condition 6 references the case when the rule references two different letters that have different attributes. The aim here is to identify the effects of these semantic similarities or differences on subject behaviour. The last condition 7 tests if the colours are obtained from a small subset of possible colours where the two chosen colours lie adjacent to each other on the colour circle. This is then compared to the colours of black and white that imply that a contrast exists between them.

Now that all the design issues of the various conditions have been reviewed, one can go on to take a look at the actual experiment that was run.

## 5.2 Experiment

An experiment was run based on the above design with two tasks placed on two pages, the first contained the selection task and the second asked subjects to write the "*negation*" of the antecedent and consequent.

### 5.2.1 Subjects

277 volunteer students from a private high school in Bahrain all of whom studied at level 10 or 11. 33 students performed condition 1, 42 students performed condition 2, 42 students performed condition 3, 52 students performed condition 4, 27 students performed condition 5, 38 students performed condition 6 and 43 students performed condition 7. The task was given to subjects during their homeroom period ahead of the start of the school day to avoid any influence they may carry from any courses taken and it was done by their own school teachers on the same day so no student could inform the others on what selections were made.

### 5.2.2 Materials

There are 7 conditions of this experiment and one of them, namely condition 3, is presented in full below:

**Condition 3:**

"*Below is depicted a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, there is a number on one of its sides and a letter on the other. Notice that numbers are white, while letters are black.*

*Also below there is a rule which applies only to the four cards. Your task is to decide which of these four cards you must turn in order to decide if the rule is true. Don't turn any unnecessary cards. Tick the cards you want to turn.*

***Rule:*** If a card has a black 'A' on one side then it must have a white '7' on the other side.

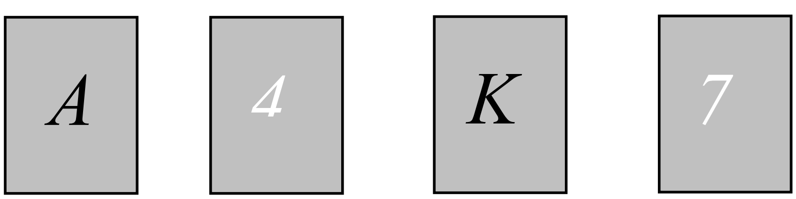



Figure 5.1: Cards shown to subjects in this task.

The second page attempts to identify the values of the negation of premises.

**In that experiment:**

If you were told that S = "*A card has a* *black A on one side*"
Then you assumed Not S = "*A card has a* ______________________________"

If you were told that S = "*A card has a* *white 7 on the other side*"
Then you assumed Not S = "*A card has a* ______________________________"

The other conditions are all similar to the above with the exceptions shown in sections 5.1.2.1 through to 5.1.2.4.

## 5.2.3 Results

This section will be composed of three main parts: a section on the replication of the original distribution, a section on the results of the selection task, followed by the results of the "*logical negatives*" task. Finally, there is a section that links the results of the "*logical negatives*" with the selection results. The order of presentation aims to reduce arising complexity by starting with the most straightforward and ending with the most complex.

### 5.2.3.1 Replication Results

The main aim of the first condition is to provide a basis with which to compare the various other results that are obtained. The results obtained from condition 1 will, therefore, be utilised as a baseline from this point forward and are shown in table 5.3. There is not significant statistical difference in any of the categories shown when tested with a Fisher Exact test on a per column basis.[10] Consequently, it is safe to assume that replication has been achieved in this condition that was run simultaneously with the others on the same target population.

| Abstract | pq | q | p | p¬q | ¬q | ¬pq | pq¬q | ¬p¬q | p¬pq | q¬q | ¬p | p¬p | p¬p¬q | ¬pq¬q | all | none | tot |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Wason** | 10 | 0 | 3 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 16 |
| **School Data** | 19 | 2 | 12 | 2 | 3 | 1 | 0 | 5 | 0 | 3 | 1 | 4 | 0 | 0 | 0 | 0 | 52 |

Table 5.3: A comparison of the replication condition with Wason's (1968) data

[10] This test will be continually used throughout this chapter because of the sparseness of the results.

### 5.2.3.2 Distribution of Subject Choices

Thus far, the best possible means of comparing subject responses to the baseline is through the comparison of the distribution of subjects' sets of choices amongst the various alternatives.

Table 5.4 displays subject distribution among the various possible choices in the main task for all conditions including the baseline condition 1. Figure 5.2 follows with the bar graph representation of the percentage values of the same data.

| Condition | pq | q | p | p]q | ]q | ]pq | pq]q | ]p]q | p]pq | q]q | ]p | p]p | p]p]q | all | total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **C1** | 19 | 2 | 12 | 2 | 3 | 1 | 0 | 5 | 0 | 3 | 1 | 4 | 0 | 0 | 52 |
| **C2** | 10 | 5 | 12 | 1 | 4 | 0 | 0 | 2 | 0 | 2 | 0 | 6 | 0 | 0 | 42 |
| **C3** | 10 | 6 | 8 | 0 | 1 | 0 | 0 | 3 | 0 | 3 | 2 | 0 | 0 | 0 | 33 |
| **C4** | 23 | 4 | 1 | 0 | 2 | 0 | 0 | 7 | 0 | 4 | 0 | 1 | 0 | 0 | 42 |
| **C5** | 7 | 3 | 4 | 3 | 0 | 0 | 0 | 1 | 0 | 4 | 1 | 3 | 0 | 1 | 27 |
| **C6** | 10 | 3 | 4 | 7 | 2 | 2 | 0 | 5 | 0 | 3 | 2 | 0 | 0 | 0 | 38 |
| **C7** | 15 | 3 | 1 | 3 | 0 | 6 | 0 | 9 | 0 | 5 | 0 | 1 | 0 | 0 | 43 |

Table 5.4: Subject card choices across the different conditions

Condition 1 is the one that replicates Wason's (1968) results as was shown in the previous section. Condition 2 replicates[11] condition 1 as well as the original Wason (1968) data. This indicates that emphasising the distinction between the values that fit the individual clauses of the rule versus those that do not does not result in any difference in behaviour.

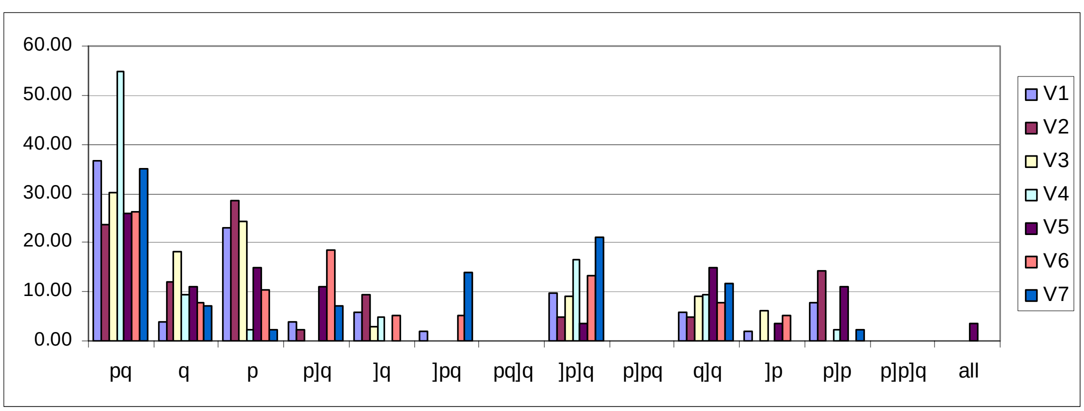



Figure 5.2: Percentages of combinations of card choices by condition

[11] The word replicated is used whenever no significant differences emerge in any of the compared categories.

Condition 3 does not replicate this baseline, because of a sole significant difference existing in Just Q selections, which indicates either that a lack of alignment with truth, or that colour marking the face and back of the cards, or both, may contribute to this result. Since condition 2 replicates the baseline while colour partitions the set and complement sets, one may assume that alignment with truth is a strong factor that may cause or hinder replication of the baseline. So perhaps condition 2 replicates because the factor is present and condition 3 does not replicate because the factor is not present. The effect could have been either just in the antecedent or just in the consequent of the rule. In fact, the results show a high selection rate in Just Q choices in condition 3 ($\underline{p} < 0.035$) when compared to the baseline.

The main difference between Version 2 and Version 3 is that the foreground rule of the latter assigns a different colour feature to the consequent than that assigned to the antecedent while the former assigns the same colour to both in the foreground rule. If this colour assignment has an effect, then the location of difference should exist either just in the antecedent or just in the consequent of the rule. In fact, results show a high selection rate in Just P choices in this version as tested with a Fisher Exact test with $p < 0.035$ when compared to the baseline to indicate that the difference is indeed localised in one of the two.

Condition 4's results are significantly different from condition 2's with a lower Just P and higher P, Q. This indicates that the common similarities that both the set and complement sets are partitioned through colour, are not sufficient to overcome their differences. The two differ in that condition 2 is aligned with truth and condition 4 is not. The differences between the two are low P selections ($\underline{p} < 0.001$) and high P,Q selections in condition 4 ($\underline{p} < 0.004$).

The same differences emerge when one compares condition 4 with condition 3. In condition 4 Just P selections are low ($\underline{p} < 0.005$) and P,Q selections are high ($\underline{p} < 0.029$). Here, the main difference in the conditions is that the order of colours is reversed in the antecedent and consequent.

Condition 4 differs from condition 2 and condition 3 in exactly the same way. When one looks at the colour marks on the cards, condition 2 is aligned with truth, while condition 3 always has one colour on one side of the card and the other colour on the other side. However, condition 4 is not aligned with the sides of the card, nor with the antecedent and consequent of the rule. The colours in this case are cross-cut so it should be expected that the differences of this condition from the other two are similar.

Now before moving onto the second group of conditions, one may recall that the baseline P, Not Q selections are at a percentage of 3.85%. Condition 5 results in a sharp rise in P, Not Q selections to a percentage of 11.11% relative to condition 3 which is 0% and ($p < 0.037$). However, it is not significantly different from the baseline. Condition 6 shows an even higher rise in P, Not Q to due the restrictions imposed by the use of colour. P, Not Q is selected with a percentage of 18.42% ($p < 0.028$) from the baseline while Just P selections are lower than the baseline ($p < 0.003$). Condition 7 results show a shift towards negative choices with a rise in Not P, Q choices with a significance of $p < 0.032$. Not P, Not Q choices are also increased but significant only when compared to condition 5 with $p < 0.003$. These choices are characteristic of imposing an alternative perspective onto the rule and this is discussed in further detail in section 5.3.

To sum up, Conditions 2, 3 and 4 explore subjects' reactions to a form of "*Information Packaging*" introduced into the task through different colours. Subjects seem to impose a "*grouping*" of the cases that make the rule true versus those that make the rule false. If, on the other hand, they are confronted with a grouping that distinguishes what the antecedent refers to from what the consequent refers to, then this results in a higher selection rate of Just Q. This is the same behaviour that was detected in the conveyor task when a temporal sequence was imposed between what is referenced by the antecedent and the consequent. In that case, it was interpreted to result from a presupposition of the event in the antecedent. Consequently, one explanation that exists for the result obtained here is that there may be an inherent semantic "*order*" to colours that implies a "*sequence*". Semantic sequences may exist in different domains such as that of numbers where they go from 1 to 2, 3, 4, 5 or letters where A precedes B, C, D, etc.

Figure 5.3 shows condition 4, in which colour groups depicted by the arrows pointing upwards do not conform to the hierarchical structure implied by the distribution amongst letter and number sets. Colour cross-cuts the sets at level 3 which may cause erroneous groupings, raising the brittleness of this condition and keeping it otherwise similar to the baseline.

This brings us to the second group of experiments that differs from the first in that colour distinguishes what is on the face versus what is on the back. Two of these conditions exhibit higher selection rates of P, Not Q which is the logically competent answer according to classical logic. There are two main features that these conditions exhibit in common that are different from the baseline condition. The first is that only a set of letters is referred to and the second is that a set of attributes of those letters is utilised to distinguish between what is on the face of the card, versus what is on the back. What this does is create an implication of semantic dependency between what is referenced by the antecedent versus what is referenced by the consequent. If the letter A appears in the antecedent then the consequent must have a letter other than A and the same applies for the colour used to describe A because in this case colour distinguishes what is on the face compared to what is on the back of the card. Consequently, there is a restrictive reading here of the conditional where the antecedent restricts the domain of possibilities for the consequent. Consider this example "*If the background is black, then the font used is white.*" Here, the restriction imposed is purely semantic because if the font used is black, then it will not be visible, so the semantic restriction may elicit a pattern of response of checking the cards when the background is black and the font is also selected as black.

Condition 7 is not similar to the above in a sense that the contrast between the black and white is lost. Consider the rule offered by Gigerenzer and Hug (1992) in the rule "*If an employee works on the weekend, then that person gets the day off during the week*." They cued the subject into the role of the employer who is checking if employees that take the day off during the week really did work during the weekend, making the expected competent responses Not P and Q. The employer's perspective is that of a person enforcing the rule versus the employee who can utilise the rule. Additionally, the consequent implies the employer will give the employee the day off, while the antecedent implies the employee will give another day of work.

Let us now go back to our example but with a variation; "*If the background is red, then the font used is orange*." The problem is transformed because now the visual problem will persist whether the rule is followed or not. However, the complementary relationship exists because red is close to orange and orange is close to red, so the task turns to that of enforcing the rule rather than utilising it. If subjects utilised it, then they would "*wish*" the font to be orange, and if they enforced it, then they would "*enforce*" that the background is red. The second is what takes place because subjects select Not P, Q and many select Not P, Not Q to confirm the change in perspective without any decreasing ambiguity in the task itself.

These results are obtained through the classical analysis methods utilised in this task and reveal that subjects do seem to impose associations that align with truth/falsity and consequently are influenced by divisions that align with antecedent/consequent or neither. However, within this experiment, another interpretive task was introduced in order to probe one of the functions that may play a factor in what selections subjects end up making in this task and the results are presented in the next section. This is followed with a discussion section that presents possible explanations for the various results.

#### 5.2.3.3 Negation Interpretation Results

The previous section analysed the results of subject choices independently of their interpretation. Two possible implications may emerge from this. The first occurs if the resulting distribution of interpretations of negations influences subjects' card choices. If so, then it is possible that experimenters have to date been misled when they were marking subject responses because they did not take into account that subjects have assumed different "*logical negatives*" than those assumed by the experimenter.

The second possibility is that subjects' interpretations do not influence subject card choices. This could have two possible causes. The first is that the interpretation is random and consequently has no effect on subject reasoning.

Results indicate that subjects do indeed make several different choices for the negation task and consequently analysis of the results has to follow a series of steps:

- The first is to define the major interpretations that subjects choose.
- The second is to analyse this data for significant differences in the distributions for particular conditions. This would help clarify if task conditions affect subject interpretations and what possible features may cause this effect.
- The third is to identify the Interpretation and conditions that are associated with the highest competent logical response according to the Interpretation assumed.
- Then a finer grained analysis of the results can be attempted in order to investigate if a relationship exists between the interpretation of the negation task and the selections subjects end up choosing, at least for card choices that have enough frequency to allow analysis.
- Last but not least, a consideration of any emerging general insight from the data can be undertaken.

- **First Step:** The first task therefore is to classify student responses as per the major classifications that they fall under.

**Competent interpretation:** This is the classical logical interpretation expected by experimenters. When S = "*A is on one side*" and T = "*7 is on the other side*", then under this interpretation the propositions of negations are given as Not S = "*K is on one side*" and Not T = "*4 is on the other side*". In other words, the logical negative is interpreted as the contrasting member from the same set of the member under consideration. This interpretation necessitates that the logically competent card selections as answers are A and 4.

> *The Classical Logical Rationale:* One may in a sense relate this negation to the given clause alone without giving any consideration to the rule. The logical negative of the "*A*" is the member of the same set of letters that is other than "*A*". In classical logic the clauses are dissociable from the rule because the conditionals are truth functional.

**Reverse Competent Interpretation:** This occurs when S = "*A is on one side*" and T = "*7 is on the other side*", and Not S = "*4 is on one side*" and Not T = "*K is on the other side*". In other words, the logical negative here is a member of the complement set that is also a contrasting member to the one that appears in the rule.

> *A Default Logical Rationale:* One may assume Stenning and van Lambalgen's (submitted) default logic semantics in order to justify the negation within this interpretation. The program completion algorithm offered by Stenning and van Lambalgen, translates the rule P → Q into (P ^ Not *ab*) ↔ Q by taking into account all the cases that would cause Q to be false and ensuring their falsity. These are represented by *ab* and if P is the cause of Q being true then *ab* must be false. Here, however, what matters is the conclusion. In other words, when subjects see a rule of the form P → Q they also assume that other propositions cannot be obtained from this rule as it only concerns the proposition Q. So, when they are asked to give the negation of P, they may search for the value that cannot be obtained by this rule. So the completed rule (P

> ^ Not *ab*) ↔ Q obtains Q and fails to obtain *cd*, so the negative of P is what P fails to obtain which is Not Q. If one rewrote the rule it would appear as (P ^ Not *ab*) ↔ (Q ^ Not *cd)*. By converse, the negative of Q refers to what will always fail to obtain Q and that is *ab* which is the other value possible for the antecedent.

**Reverse Rule Interpretation:** Under this interpretation the logical negative of the member set of letters mentioned in the rule is assumed to be the member of the complement set mentioned in the rule. The "*if*" instruction is regarded as relating "*logical negatives*" of each other such that S = "*A is on one side*" and T = "7 *is on the other side*", and Not S = "7 *is on one side*" and Not T = "*A is on the other side*".

> *A Default Logical Rationale:* This may seem the most confusing interpretation of all and by all means it is the most popular. If one may recall, that a rule of the form P → Q does not exclude propositions other than P possibly E that may also cause Q to be true in a different rule that is E → Q. So if one is to look for a conclusion without having P, then the conclusion obtained by E is the negative of P because it is obtained through the use of another rule. Therefore, here Q results. If, however, Q is not obtained, then it is also possible for there to exist another premise that implies that P → W that is not explicitly given here where W has nothing to do with Q. Notice that this too is a rule that may exist independently of the given rule. In this case the negative of Q is represented by P because it can obtain a value other than Q.

| | P = A | Not P = | Q = 7 | Not Q = |
|---|---|---|---|---|
| Competent Interpretations | | K | | 4 |
| Reverse Competent Interpretation | | 4 | | K |
| Reverse Rule Interpretation | | 7 | | A |

Table 5.5: A summary of subject responses associated with the interpretation made.

These results indicate that one may question the basis on which the current expectations of negatives are made. They may indeed be logical and reflect the same set of interpretations subjects impose on the rule.

- **Second Step:** The second step therefore is to classify student responses as per the definitions given.

The interpretations that did not fit into these categories were classified as "*other choices*". This latter category therefore differs because it spans a number of possible choices, which caused some difficulty in untangling due to the sparseness of the data as well as the difficulty of defining categories that do not overlap with the main three categories or with each other. For example, the largest category following the above three, has far lower frequencies of selection than those given above. The highest of these is when subjects select "*the opposite colour of Q*" as Not P and "*the opposite colour of P*" as Not Q; the frequencies here are for conditions 1 to 7 as follows: 0,3,3,2,3,1,0. Another issue that arises here is that when subjects select the negative, they may end up writing a value that is not one of the four cards. One may wonder if the cause is that a restrictive reading is imposed onto the tasks in conditions 2,3 and 4 because subjects are only shown 4 out of the 8 possibilities. First of all, the other conditions such as condition 5 exhibit similar behaviour even though the suspected cause is not present while condition 3 exhibits a low percentage in this category even though the suspected cause is present. This leaves another possible explanation, which is that subjects are relaxing the restriction of the background rule, possibly due to the vagueness of the question such that they assume a larger domain of possible choices than is present in the task.

The full set of data is shown in table 5.6 displaying the number of students who made that selection next to the percentage of the total number of students who performed that condition.

| | **Total** | **Competent** | **Reverse Competent** | **Reverse Rule** | **Other choices** |
|---|---|---|---|---|---|
| **Condition 1** | 52 | 8 (15.4%) | 7 (13.5) | 24 (46.2%) | 13 (25%) |
| **Condition 2** | 42 | 5 (11.9%) | 4 (9.5%) | 16(38.1%) | 17 (40.5%) |
| **Condition 3** | 33 | 8 (24.2%) | 5 (15.2%) | 12(36.4%) | 8 (24.2%) |
| **Condition 4** | 42 | 7 (16.6%) | 11 (26.2%) | 15 (35.7%) | 9 (21.4%) |
| **Condition 5** | 27 | 1 (3.7%) | 1 (3.7%) | 13 (48.1) | 12 (44.4%) |
| **Condition 6** | 38 | 6 (15.8%) | 7 (18.4) | 19 (50%) | 6 (15.8%) |
| **Condition 7** | 43 | 5 (11.6%) | 4 (9.3%) | 26 (60.1%) | 8 (18.6%) |

Table 5.6: Subject distribution pattern across the three major interpretations

We can add the total number of students who make the first three interpretations into a group which is made up of the following percentages according to the condition number: C1:75%, C2: 59.5%, C3: 75%, C4:78.6%, C5:55.6%, C6:84.2%, C7:81.4%. It is immediately apparent that the division between the three major interpretations and the "*other choices*" category is uniform across the different conditions. However, conditions 2 and 5 in particular stand out with the lowest percentages of 59.5% and 55.6% while all others are in their seventies and eighties. Running a Chi Square test on the difference between the major interpretations and "*other choices*" reveals a significance level of $p < 0.042$ that indicates that conditions 2 and 5 both differ in the frequency of the various interpretations of negation. In condition 2, both quantifiers "*one side*" and "*other side*" reference values that are white, which implies that the division is between what is in the rule and what is not in the rule. Condition 5, on the other hand, referenced the same letter by both quantifiers "*one side*" and "*other side*" namely letter A.

| **Interpretation of Negation** | **Selection Task Conditions** | | |
|---|---|---|---|
| | **Condition 2** | **Condition 5** | **Condition 7** |
| **Competent Interpretation** | | | |
| **Reverse Competent** | Less than C4 $p < 0.043$ | Less than C4 $p < 0.014$ | Less than C4 $p < 0.038$ |
| **Other Choices** | More than C4 $p < 0.049$ | More than C4 $p < 0.040$ | |
| | More than C6 $p < 0.013$ | More than C6 $p < 0.012$ | |
| | More than C7 $p < 0.024$ | More than C7 $p < 0.020$ | |
| **Reverse Rule Interpretation** | | | More than C3 $p < 0.032$ |
| | | | More than C4 $p < 0.019$ |

Table 5.7: Comparison between selection task conditions 2, 5 and 7 on the frequency of different interpretations of negation they evoke; only significant differences between them and other conditions made per interpretation type are displayed.

Table 5.7 clarifies the similarities between conditions 2 and 5 and differences from other conditions. Their behaviour is identical even though condition 2 has 8 possible cases of which only 4 are shown while condition 5 has 4 all of which are shown. Additionally, condition 2 has both a set of letters and numbers while condition 5 has only a set of letters. Another difference is that condition 2 implements "*information* *packaging*" onto the task, while condition 5 utilises colour to distinguish what is on the face of the card versus what is on the back.

The identical behaviour is obviously not caused by any of the above differences. These two conditions have one basic similarity; condition 2 has the same identical attribute colour associated with the letter in the antecedent and the number in the consequent, while condition 5 has the same letter value which is "*A*" in both the antecedent and consequent. One may refer to the ID principle here (Fauconnier, 1994) as it implies that an attribute could be utilised to refer to whatever this attribute defines. For example, one may say here "*The German girl knows how to solve this problem,*" rather than say "*Anna knows how to solve this problem.*" In the first sentence, the attribute that the girl is German is taken as sufficient to identify the girl without necessarily giving her name. Similarly in condition 2 the attribute of a "*white colour*" is taken as an identifier while in condition 5 the same name "*A*" is utilised to relate the antecedent to the consequent.

Condition 7 also exhibits some traits distinguishing it from the others, as it has two colours that lie adjacent to each other on the colour circle. The same influence of perspective change already discussed in section 5.2.4.1 seems to be evident here. What seems to be happening in condition 7 is that one set of "*reds*" when coupled with one set of letters restricts the possible range of "*logical negatives*" such that 60.1% of subjects make the Reverse Rule Interpretation. Although this condition does not have a unity of the letter as in condition 5 or unity of the colour as in condition 2, it does imply proximity by introducing a small subset of what may be a large domain of colours. White and black are considered contrasts that have all the other colours in between their intensity levels.

However, at this stage one should consider the implication or "*meaning*" of these various interpretations of negation and how illuminating they will be when considering subject behaviour in the task. The competent interpretation of negation shows that subjects isolate the clauses from the structure of the rule in order to perform the negation. The reverse competent and the reverse rule interpretations, imply that subjects are utilising the rule as well as the clauses within it as a means to arrive at their negative. This highlights a problem that seems inherent to descriptive selection tasks, which is that of interpretations.

First of all, the Reverse Competent interpretation exhibits a constant awareness of other factors that may influence the conclusions that are not immediately clear in the question. Taking these "*other*" possible influences causes them to give this negative. By contrast, the most common interpretation of negation is the Reverse Rule interpretation. This results from taking the given rule to be part of a larger world that has many different rules of which this is only one. Perhaps in this case, the rule is taken as one out of a sample of possible rules and if one may recall, Stenning and van Lambalgen (2004) found that some subjects take the cards to be a sample out of a larger set of cards.

So far, one can be certain from the above results that the condition can affect subject interpretations which makes one wonder if there is any relationship established between the interpretation assumed and the choices made.

#### 5.2.3.4 Considering Subject Selections from the Perspective of their Negation Interpretations

One must take a look at the data in order to identify if it is possible to perform various types of analysis. Although it would be wonderful to run a full analysis on the data for all possible interpretations and conditions, this is not possible here due to sparseness. Consequently, subject distributions amongst the various choices are reported for the Reverse Rule Interpretation. However, here since Not P and Not Q are not interpreted as researchers would expect, the data is only reported in terms of raw data with the letters P, Q, R and S which correspond to the value A, 7, K, 4 in the first four conditions and analogously for the others.

| | | **PQ** | **Q** | **P** | **PS** | **S** | **RQ** | **RS** | **SQ** | **R** | **PR** | **all** | **Total** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| reverse rule | C 1 | 13 | 0 | 5 | 1 | | | 2 | 3 | | | | 24 |
| reverse rule | C 2 | 4 | 2 | 4 | | 3 | | 1 | | | 2 | | 15 |
| reverse rule | C 3 | 4 | 2 | 2 | | 1 | | 1 | 1 | 1 | | | 12 |
| reverse rule | C 4 | 12 | 2 | 0 | 0 | | | | 1 | | | | 15 |
| reverse rule | C 5 | 2 | 0 | 3 | 1 | | | 1 | 3 | | 2 | | 12 |
| reverse rule | C 6 | 7 | 0 | 3 | | 1 | 2 | | 1 | 1 | 4 | | 19 |
| reverse rule | C 7 | 13 | 1 | 0 | 1 | | 4 | 6 | 1 | | | | 26 |

Table 5.8: Frequencies of card choices made by subjects by interpretation and condition

Yet again the columns are sparse so it would be wise to only consider the *Reverse Rule Interpretation*. The frequencies will be considered for P, Q selections also because of the high frequencies in this category in an attempt to associate interpretations with subject card choices. However, before doing that, it is possible to estimate just how influential this task is when it comes to predicting P, Q selections by comparing the frequency of selecting P, Q in comparison to the frequency of selecting P, Q in other interpretations. This is done in table 5.9.

| | **Non Reverse Rule Interpretations** | | **Reverse Rule Interpretations** | |
|---|---|---|---|---|
| | P,Q selections in General Case | Non P,Q selections in General Case | P,Q selections in Reverse Rule Interpretation | Non P,Q selections in Reverse Rule Interpretation |
| **Condition 1** | **6** | **25** | **13** | **8** |
| **Condition 2** | 6 | 21 | 4 | 11 |
| **Condition 3** | 6 | 20 | 4 | 3 |
| **Condition 4** | 11 | 8 | 12 | 11 |
| **Condition 5** | 5 | 10 | 2 | 10 |
| **Condition 6** | 3 | 16 | 7 | 12 |
| **Condition 7** | **2** | **15** | **13** | **13** |

Table 5.9: Frequency of selections as a function of interpretation of negation; P,Q and non P,Q selections and RR[12] and non RR interpretations by condition.

Table 5.9 shows this comparison where results show that conditions 1 and 7 show significant differences between P, Q selections in the general case versus non P, Q selections as compared to P, Q selections in the Reverse Rule Interpretation, versus non P, Q selections with $\underline{p} < 0.002$ and $\underline{p} < 0.010$ when tested with a Fisher Exact Test. This implies that the Reverse Rule Interpretation is a major contributor to the P, Q selections made in at least these two conditions which is sufficient to warrant consideration.

Under this interpretation subjects may be assuming that there is more to the question than meets the eye. Given the rule, "*if P then Q*", they seem to assume that P may imply consequences other than Q and that antecedents other than P may imply Q. This reflects a concern with what may be semantically acceptable but not necessary directly deduced from the rule versus what may never be allowed by any generalisation of the rule.

## 5.3 General Conclusion

The previous chapter offered a question scenario that resolved the semantic anaphora imposed in the selection task replacing it with a temporal sequence that is enforced. Results showed that a constant anaphoric interpretation is elicited by this scenario with a high selection rate of Just Q. A justification for this was offered based on the temporal context of the task.

[12] Reverse Rule Interpretation.

In this chapter, the conditions tested fell into two main groups. The first examines the AK47 task through the use of "*information packaging*" techniques. Distinctions are emphasised in order to identify whether they are aligned with or whether they intersect with distinctions that are imposed by subjects onto the task. Results showed that emphasising the distinction between what is referred to by the rule and what is not aligns with subject distinctions and does not alter behaviour from the baseline. By contrast, distinguishing what is referred to by the antecedent from what is referred to by the consequent results in high Just Q selections.

This indicates that there could be a link between this result and that obtained in chapter 4 raises a question of whether they have similar causes, but one should not forget that they may also have completely different causes. The similarities between the two rest in the fact that in chapter 4 the distinction between what the antecedent and consequent refer to is enforced through time separation as they occur in a preset sequence. In condition 3, here the distinction is imposed through the utilisation of different colours for what is referred to by the antecedent and what is referred to by the consequent.

A conditional is a construct that relates the antecedent to the consequent in a precise uni-directional fashion which implies that the relationship describes a distinction between the antecedent and the consequent that is deduced from it. On the other hand, cards can have "*one side*" refer to both the "*face*" and "*back*" of the cards (Stenning and van Lambalgen, 2004). The relationship the sides of the cards have to each other does not map onto that between the antecedent and consequent. In the rule, there is a distinction, while on the cards there is no distinction. Consequently, it should not be surprising that emphasising this particular distinction should result in a constant anaphoric interpretation because it imposes onto the cards, the distinction that already exists in the rule. One may wonder why high Just Q choices rather than high Just P choices result if there is a constant reading? A possible answer to this is that the distinction emphasised here is between what is referred to by the antecedent and consequent, so it reflects the "*consequence*" relation between them as implied by the conditional, which it in fact does, as the consequent is what is selected.

The second group of experiments emphasise the effect of these distinctions because the same order of colours in condition 3 is retained. These were designed as a form of control experiment to identify if the 8 possibilities of which only 4 are shown in the first group have any effect on subject behaviour. Reducing the possibilities to 4 meant that the hierarchical structure of the referents was altered. Colour was no longer imposed as an information packaging tool. In conditions 5 and 6 set membership as well as colour attributes ensure that subjects understand that both the antecedent and consequent referents come from the same domain. At the same time, colour distinguishes what is on the face from what is on the back. These resulted in relatively high selections of P, Not Q.

This questions the formulation of the Wason (1968) selection task as the hierarchy of referents imposes semantic distinctions that lack a unified domain that relates letters to numbers with anything other than the rule placing them on the sides of the same card. This context is unable to elicit the desired P, Not Q responses because it does not evoke the interpretation that would guide subjects towards these choices. At least one of the possible requirements is a unified domain that contains both what the antecedent and consequent refer to.

It should, therefore, not be surprising that subjects revert to the rule when asked to perform the negation task, as this rule exists to relate the letter to the number by placing them on two sides of the same card. It is not surprising either that subjects carry along their "*assumptions*" and impose them onto the negation task. In this task, they seek to negate by searching for what else can be deduced or what cannot be deduced. The line that separates a value from its negative is altered because the rule is what relates the referent of the antecedent to that of the consequent.

Condition 7 tests this by representing two consecutive letters and two similar colours in a rule which results in a very high selection rate of the Reverse Rule Interpretation. This implies that when the referents are similar to each other or consecutive to each other, subjects seek negation by searching for what is not in the rule, as opposed to making a more diverse set of responses if there is contrast as with black and white colours.

This leads us to wonder about the nature of the relationship between these negations and subject interpretations. Unfortunately, due to the sparseness of the data, this could only be tested with respect to P, Q selections that seem to correlate well with the Reverse Rule Interpretations. One of the possible logical justifications of this interpretation is when subjects ask what P can obtain other than Q. In other words, "*If P then Not Q*", which is characteristic of the "*logic of false*" as defined by Stenning and van Lambalgen (2004), is offered as an explanation for high selections of P, Q which correlates well here with the Reverse Rule Interpretation.

To sum up, the interpretations subjects impose are semantic because they are affected by semantic alterations in the task. In general, there is evidence for a distinction between the rule referents and other values not referred to by the rule. There is also evidence for the lack of or the existence of a weak distinction present between what is referred to by the antecedent and what is referred to by the consequent, because emphasising it alters behaviour in the task. The hierarchy of the values referred to by the rule also affects performance, as subjects are prone to make more P, Not Q selections if there is more unity of the domain from which the referents are extracted rather than selecting them from completely distinct domains that are not related except by appearing on the sides of the same card. Last but not least, the negation task reveals that experimenter assumptions to date have been misguided because even when performing negations, subjects exhibit the same interpretations they exhibit in the task. Here, a strong correlation is shown between the Reverse Rule Interpretation and P, Q choices which is explained by a "*logic of false*" assumption that is in turn implied by the Reverse Rule Interpretation.

# Chapter 6

# In Search of Surface Feature Classification Traits

The last two chapters ran a deep semantic analysis of the relationship between the terms and what they reference. However, one should never overlook that each of these terms exists in a context which affects its interpretation. The word "*letter*" for example, may exist in the context "*the letter N*" or it may exist in the context "*the letter has a stamp on it*" but is not likely to exist in the context "*the letter wrote a book*". Some may claim that the words surrounding the word "*letter*" play a role in determining its meaning (Landauer and Dumais, 1997) while others argue against that (Glenberg and Robertson, 2000). One thing for certain is that words do not live in isolation of their contexts and in the case of the word "*letter*" for example, the words appearing in that context do seem to have an effect on the intended interpretation of the word. Consequently, an analysis of surface co-occurrence features seems essential. In other words, this chapter will seek through co-occurrence a means to compare rule antecedents to consequents in the various rules listed in Appendix A. This will require the utilization of a computational model that studies co-occurrence trends of the words with the aim of finding a distinction between deontic and descriptive rules; namely Latent Semantic Analysis or LSA (Landaur and Dumais, 1997). To date, the distinction between these two types of rules could only be achieved through subjecting rules to the criteria described in chapter 3, and consequently it has to be done manually. So finding a distinction computationally should resolve whatever doubts that may remain that there are two distinct categories. However, LSA has a central problem that may cause the interpretation of the results to be more complicated. Basically, it ignores the order of the words in context and consequently deals with contexts as bags of words. A novel neural network self organizing map is offered to resolve this problem and this network is tested with the same rules. The network classifies the rules in various regions of tensor space while only comparing words to other words in comparable positions. The results therefore should show how significant the word positions are in influencing the classification into the two groups. This should be more informative in the sense that it will give a better description of a descriptive rule's surface format as compared to a deontic rule's surface format with respect to the co-occurrence and positional behaviour of the words they contain.

## 6.1 A Little History of LSA

The data representation utilised here is obtained from an online site that retrieves information from a model called Latent Semantic Analysis (Landauer and Dumais, 1997). With the immense number of electronic documents in the world, scientists felt the need to develop a computational system that is capable of classifying them and auto retrieving them in an automatic fashion.

For example, if a person is running the search keys in the words "*letter*" and "*admission*" then this person is also likely to be searching for the combination "*letter*" and "*acceptance*" without necessarily typing the synonym. Another issue is that the sense of the word "*letter*" here is from a different context than the one in which the sentence "*the letter is N*" may exist. Our search employee will not like retrieving a number of documents with information about the letter N. This second sense of the word letter is called a homograph meaning of the same word.

These frequently occurring problems led researchers to develop methods to extract information from higher order associations in order to make the search more informed (Deerwester, et al., 1990). Latent Semantic Analysis was developed for indexing by using these associations to resolve problems with homographs and synonyms. Following its development, comparisons were made with children's learning rates to find that it seems to explain how they learn new words from a context.

One should not jump to the conclusion that this model reflects meaning as there is still doubt about that. However, one should take the time to ponder what LSA is actually doing. Only through careful speculation can one make sense of what is going on, enough to understand what these numbers actually reflect. A description of the mathematics behind the model is given in Appendix B to allow the reader to concentrate on the motivations, advantages and disadvantages of the model.

A central issue is that LSA takes contextual data from an encyclopaedia in contexts of 2000 words each, so there is no dispute that the amount of data is immense. Additionally, the resources used to build it are based on reading materials that span the sciences of the world, so their relationship with the concepts that exist in the world cannot be overlooked. However, what exactly does it do with all that? LSA does not read the words, so the word "*ugly*" when used as input is the same to LSA as the word "*beautiful*", just as the word "*dead*" versus the word "*alive*". Consequently, it is extremely important to inspect what LSA is actually doing with words in order to arrive at its similarity figures.

It starts off by calculating how often each word appears in each context. This is like saying that I will analyse a company's activities by partitioning them into what occurs in each project. A project here represents a context. The model then tries to find out if those same people work in different projects during the same period, so it calculates for each person how many projects they are involved in. Then it tries to approximate the employee/project relationship such that it generalises what is going on, into several main areas of specialisation for that company. These areas of specialisation are in fact of a lower number than the projects people actually work on.

It should be clear from the analogy that this system is very far from attaining deep meaning. However, it should also be clear that this generalisation retains key bits of information that could be easier to analyse than if all is retained or if there were none at all. For example, it would have an approximate figure of how frequently word X and word Y appear in similar contexts. Table 6.1 exhibits a comparison of some words that may appear in selection task rules.

| | Person | Overnight | A | N |
|---|---|---|---|---|
| Person | 1 | 0.10 | 0.36 | 0.03 |
| Overnight | 0.10 | 1 | 0.40 | 0.10 |
| A | 0.36 | 0.40 | 1 | 0.26 |
| N | 0.03 | 0.10 | 0.26 | 1 |

Table 6.1: Selected pair-wise word co-occurrence similarity figures as obtained from LSA

The similarity factor between the letters A and N is 0.26 while that between "*Person*" and "*Overnight*" is 0.10. The first two are from a descriptive or abstract rule while the second two are from a deontic rule. Deontic rules have never to date been compared to descriptive rules with respect to how they look and the words they actually contain. It is quite possible that the distinction between the two types of rules could exist on the surface level of the rules themselves and not just at the deep semantic level already discussed. In table 6.1, the letter "*A*" may represent a function word if it appears in the form "*If there is a ..*". Function words are words that carry more syntactic that semantic meaning and these include prepositions, articles, auxiliary verbs, pronouns and such (Smith and Witten, 1993). This type of word carries little meaning on its own due to the large diverse set of words that can follow it like "*a*" in the examples "*a car*", "*a game*", "*a house*", "*a letter*" which describe both activities as well as objects using the same article. On the other hand, the word "*person*" is very different because it can only refer to a human being so the class it refers to is described as an open lexical class. This is the class of content words whose meaning is more concrete and which includes nouns, adjectives, verbs and the like (Smith and Witten, 1993).

If we now consider the table again then the similarity measure between "*A*" and "*person*" is 0.36, while between "*A*" and "*overnight*" it is 0.40 which implies that function words appear close to a large range of lexically open class words that include "*person*" and "*overnight*". Consequently, it is worth considering if this semantic distinction between the two types of words which clearly affects their behaviour in context may be sufficient to reflect a distinction between the two types of rules.

## 6.2 Comparing Rule Surface Features with LSA

Perhaps the first issue that comes to mind is what should be compared versus what should not be compared. First of all, there are the rules that were used in the various experiments listed in Appendix A. This is a definite candidate that will retain its ground for now as we look for others.

Second, there are the contexts used by researchers when giving students the tasks. Unfortunately, not all researchers reported the exact text of the presented materials used in their experiments. Another issue that arose is that the lengths of the materials varied from several pages to a sentence or two which made the following stage of formalising it for the neural network practically impossible. For these reasons a comparison of contexts became impossible.

This leaves us with the rules, and how much information co-occurrence statistics can tell us about how similar the rules are. Let us take a look at a few examples to identify the difficulties that may arise. An example of a descriptive rule is; "*If the letter is N, then the number is 3*" (Manktelow and Evans, 1979) while an example of a deontic rule is "*If an employee works on the weekend, then that person gets a day off during the week.*" The similarity factor between these two is 0.56 while the similarity measure between the first of the two and the rule "*If the tablecloth is brown then the wall is white*" is 0.51. This evidently is one of LSA's problems as the similarity of the If .. Then .. construct seems so influential as a classifier that it fails to distinguish between deontic and descriptive rules. This also informs us about the type of differences that LSA utilises to distinguish sentences, which is simply that the use of the same words is highly influential. The repetition of the words "*If*" and "*then*" placed almost all rules in the same category. A more extensive test of the full list of deontic rules versus descriptive rules was run to reveal that LSA, as expected, finds them all highly similar.

This does not imply that LSA is unable to help in analysing rule surface structure because all it was doing was to classify them as rules, which they all are. It is not as much an error as it is the wrong level of analysis. The previous two chapters should be sufficient to point out that even while the conditional structure remains the same, subject behaviour is influenced by differences between the antecedent and consequent clauses. So this raises another possible avenue to follow where one can actually compare the antecedent to consequent in each of the deontic and descriptive rules to see if there are any obvious trends.

### 6.2.1 Co-occurrence Distances Between Antecedents and Consequents

The target of this study is to search for surface co-occurrence distances between antecedents and consequents and to utilise results to further understand the data provided by LSA. In order to do this the full list of rules shown in Appendix A was first categorised to be either deontic or descriptive. This categorisation was performed strictly in accordance with the definitions given in Chapter 3 and all rules that were difficult to categorise or seemed controversial were excluded.

Then the antecedents and consequents of each rule were extracted by only removing the words that construct the conditional which are "*if*", "*then*" and "*only if*". Following this each rule was tested individually on the LSA site in order to obtain a co-occurrence figure for the distance it gives between the antecedent and consequent. Results are shown in table 6.2.

| Deontic Rules | Premise Proximity In LSA | Descriptive Rules | Premise Proximity In LSA |
|---|---|---|---|
| **R2** | 0.58 | **R1** | 0.33 |
| **R4** | 0.58 | **R3** | 0.33 |
| **R5** | 0.58 | **R7** | 0.66 |
| **R6** | 0.59 | **R9** | 0.47 |
| **R13** | 0.58 | **R10** | 0.47 |
| **R14** | 0.58 | **R11** | 0.72 |
| **R15** | 0.24 | **R12** | 0.72 |
| **R16** | 0.24 | **R17** | 0.66 |
| **R21** | 0.51 | **R18** | 0.58 |
| **R22** | 0.26 | **R19** | 0.84 |
| **R38** | 0.33 | **R20** | 0.64 |
| **R42** | 0.33 | **R25** | 0.72 |
| **R46** | 0.33 | **R26** | 0.58 |
| **R56** | 0.5 | **R27** | 0.64 |
| **R57** | 0.63 | **R28** | 0.72 |
| **R58** | 0.5 | **R29** | 0.58 |
| **R59** | 0.63 | **R30** | 0.64 |
| **R60** | 0.64 | **R31** | 0.72 |
| **R61** | 0.57 | **R32** | 0.58 |
| **R62** | 0.63 | **R33** | 0.64 |
| **R63** | 0.57 | **R34** | 0.58 |
| **R66** | 0.38 | **R35** | 0.58 |
| **R67** | 0.38 | **R36** | 0.58 |
| **R68** | 0.49 | **R74** | 0.61 |
| **R69** | 0.49 | **R75** | 0.61 |
| **R70** | 0.61 | | |
| **R71** | 0.61 | | |
| **R72** | 0.39 | | |
| **R73** | 0.39 | | |
| **R76** | 0.57 | | |
| **R77** | 0.57 | | |
| **R78** | 0.26 | | |
| **R79** | 0.26 | | |
| | | | |
| **Mean** | **0.4788** | | 0.6073 |
| **St Dev.** | **0.1355** | | 0.1139 |

Table 6.2: LSA co-occurrence similarity results when comparing the antecedent and consequent of each rule to each other.

The data displayed shows that LSA exhibits a clear distinction between the two groups with a mean difference of 0.1285 which is not a slight difference at 21% difference of the deontic rules.[13]

However, there does seem to be an overlap with respect to some rules including from the Deontic side; "*If you eat duiker meat, then you have found an ostrich eggshell*", "*If someone stays overnight in the cabin, then that person must bring along a bundle of wood from the valley*" and "*If a previous employee gets a pension from a firm, then that person must have worked for the firm at least ten years*".

Additionally, there is overlap from the descriptive side in the rule "*If a letter is sealed then it must carry a 20-cent stamp*".

### 6.2.2 Discussion

The aim of this level of analysis is to identify the causes of the distinctions found by LSA. In order to be able to comprehend what is going on, a word to word comparison within one of the rules that caused the overlap is considered in table 6.3.

| | that | person | must | bring | along | a | bundle | of | wood | from | the | valley |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| someone | 0.49 | 0.58 | 0.25 | 0.31 | 0.23 | 0.49 | 0.13 | 0.40 | 0.10 | 0.38 | 0.44 | 0.04 |
| stays | 0.31 | 0.12 | 0.17 | 0.24 | 0.18 | 0.35 | 0.10 | 0.31 | 0.15 | 0.34 | 0.34 | 0.08 |
| overnight | 0.38 | 0.10 | 0.14 | 0.32 | 0.20 | 0.40 | 0.11 | 0.37 | 0.05 | 0.38 | 0.39 | 0.11 |
| in | **0.96** | 0.32 | 0.55 | 0.61 | 0.51 | **0.98** | 0.32 | **0.99** | 0.26 | 0.95 | **0.99** | 0.24 |
| the | **0.96** | 0.32 | 0.56 | 0.62 | 0.54 | **0.99** | 0.36 | **0.99** | 0.27 | 0.95 | **1.00** | 0.25 |
| cabin | 0.28 | 0.05 | 0.08 | 0.24 | 0.28 | 0.32 | 0.28 | 0.28 | 0.35 | 0.29 | 0.32 | 0.22 |

Table 6.3: Word to word co-occurrence similarity figures of a deontic rule.

The words in the column to the left are the ones in the antecedent while those at the top of the table are those that form the consequent of the rule. The highest similarities are oddly reported between articles "*in*", "*a*" and "*the*" so the fact that this rule has so many may be the cause for a high similarity measure. A descriptive rule is shown in table 6.4.

[13] Deontic rules' figure – descriptive rules' figure to give a result that is multiplied by 100 and divided by the Deontic rules' figure

| Document | it | has | a | \'4\' | on | the | other | side. |
|---|---|---|---|---|---|---|---|---|
| **a** | 0.95 | 0.68 | 1.00 | 0.32 | 0.95 | 0.99 | 0.85 | 0.39 |
| **card** | **0.11** | **0.11** | **0.13** | **0.07** | **0.17** | **0.12** | **0.11** | **0.06** |
| **has** | 0.65 | 1.00 | 0.68 | 0.26 | 0.61 | 0.68 | 0.68 | 0.22 |
| **an** | 0.82 | 0.69 | 0.91 | 0.32 | 0.83 | 0.89 | 0.79 | 0.32 |
| **\'A\'** | 0.95 | 0.68 | 1.00 | 0.32 | 0.95 | 0.99 | 0.85 | 0.39 |
| **on** | **0.93** | **0.61** | **0.95** | **0.31** | **1.00** | **0.96** | **0.80** | **0.46** |
| **one** | 0.91 | 0.67 | 0.95 | 0.30 | 0.90 | 0.94 | 0.86 | 0.42 |
| **side** | 0.40 | 0.22 | 0.39 | 0.13 | 0.46 | 0.40 | 0.38 | 1.00 |

Table 6.4: Word to word co-occurrence similarity measure of a descriptive rule.

The words that appear in many different contexts are usually function words such as "*the*" and "*in*" while those that are more restricted in the contexts they appear in are mostly open lexical class words like "*bird*" which has a referential semantic function. Analysis of the results seems to imply a possibility that it is this distinction that is picked up by LSA between the two categories.

It should be clear at this stage that LSA which investigates the frequency of word use seems to categorise the rule according to two main groups of words. The word "*on*" is a function word that appears in many different contexts so it shows a high similarity measure to almost all words in the consequent. The other group is represented by the word "*card*" which is an open lexical class because it has a semantic value by reference to a class of objects which is described by the word "*card*". This word reflects a low similarity measure to all words in the consequent.

Perhaps one way to comprehend what is going on is to go make a word to word comparison of all words that appear in the list of rules in Appendix A. Words were collected and compared. Each word's similarity measure was obtained from LSA's online site in terms of all other words including itself. Then the average similarity measure of each word was calculated by averaging all these figures. This number now reflects how "*general*" this word's occurrence is because the higher it is the more likely it is for this word to appear in among most of the other words and the lower it is the more likely it is for this word to only appear along with only a few of the others. The words' "*generality*" measures had a maximum value of 0.45 so in order to isolate only the most general terms, those that scored above 0.25 are listed as follows: A, IF, THE, THEN, ON, IS, TO, IT, AN, HAS, ONE, HAVE, OTHER, OF, BE, BY, THAT, HE, ONLY, AT, IN, ANY, FOR, YOU, AND, EACH, FROM, LEAST, OR, OVER, BEFORE, BRING, CLUB'S, GET, HAD, HIS, LIKE, SMALL-TIME, TAKE, THEY, TWO, WAS, WITH.

This list includes 8 adjectives that can be classified as per their use in the rules, 8 verbs, 2 nouns and 25 function words. Some of the adjectives can be considered function words in other contexts which may have affected LSA's similarity measure. LSA is utilised to obtain both the generality measure as well as LSA's similarity measure so if there is a generalisation to interpret the words that can take both roles as function words then the same applies to both measures causing the conclusion to be valid even if the error exists. In short, the similarity measure given by LSA is relatively high when comparing words that carry little lexical meaning and appear in many different contexts to other words that appear in the texts.

Thus, the list of words given shows that the type of word that carries little semantic meaning seems to affect the overall similarity measure by raising it and its frequent existence in descriptive rules as compared to deontic rules, and may be the cause of a distinction at the surface level of analysis. Rules that are part of the overlap that occurs in the categorisation are those that do not conform to this surface distinction as was shown in table 6.3 because they contain a number of function words that do not fit their own category.

Although these results show that the distinction that is captured may indeed be a distinction in the frequency of function words appearing in one of the two categories versus the other, it offers no reason to exclude other possible culprits of the categorisation that goes on. Function words in general exist to provide grammatical relationships with other words in the sentence or to express the mood of the speaker. Consequently, their position in the sentence is crucial to the role they play in the sentence.

Additionally, function words are also acquired late during language learning which led to the assumption that they involve more abstract grammatical knowledge. This in turn led to the assumption that they play a structural role that in turn aids in categorising open class lexical words (Smith and Witten, 1993). This attention to the location of function words with respect to open lexical words imposes a need to identify if the same distinction persists if the order of the words is enforced as the same.

LSA is frequently criticised for neglecting order (Wiemer-Hastings, 2000) such that the two following sentences appear identical in meaning representation according to LSA.

***Mark killed the tiger vs The tiger killed Mark***

It is accepted without saying that the death of a human is never to be compared with the death of an animal so one may wonder if this lack of "*order*" may affect the distinctions it seems to find. One problem is therefore raised, and it is simply how to add order to LSA in order to identify if function words are the primary cause of the distinction because they have strong structural roles.

### 6.2.3 Ways to Add Order

This importance of adding word order was recognised by Wiemer-Hastings (2000) who investigated the effects of the neglected syntactical information in a sentence and attempted to incorporate it into the LSA framework. He separated the sentences into atomic clauses or propositions and then segmented them by hand to break them into strings composed of subject noun phrase, verb and object noun phrase. Antecedents were used to resolve pronouns and conjunctions were dealt with by distributing the arguments. Then he attempted to evaluate the similarity of this presentation using a variety of measures. Results showed that the best approach to combine the similarities of the sentence parts is non-linear and even that was not as close to human judgment as LSA. Wiemer-Hastings and Zipitria (2001) then went on to a further test, incorporating syntax through two methods. The first was to tag the words used for the training corpus at 100, 200, 300 and 400 dimensions and this did not produce any favourable results. Then they tried a structured LSA or SLSA where they broke up sentences into parts as was described above and used that as training material to find results that correlate slightly better than LSA.

Their partial success is a stronger indicator that this course should be pursued further, except that it may the case that adding part-of-speech tags is not the right way to go about it. Since LSA does not read the words by identifying their syntactic role, nor semantic role, it may be the case that the way to add order is not through adding any "*role*" identifying features. Instead, the addition should treat words the exact same way as LSA treats them, which is simply as vacant symbols whose behaviour is analysed when in the proximity of other symbols. Since LSA does this though a vector representation, the only solution is to raise the representation one level to a tensor representation.

## 6.3 A Brief Meeting with Tensors

Unfortunately, the name *tensor* seems to be constantly linked with Riemann geometry and the theory of Relativity. This strong relationship seems to invoke an impression that whatever is linked to this type of mathematics must be incomprehensible. The truth of the matter lies in the fact that any topic that is regarded from the outside, seems to be incomprehensible, so the best way to understand tensors is to visit some of the tensor representations that we have in our world. We can begin by visiting a beautiful Swiss Alpine mountain, for example, that can be seen from the top of another mountain. What one can see is represented in three dimensions, length, breadth and breathtaking height. This is a tensor of the lowest rank.

Prior work in representing text through tensors was done by Smolensky (1990). He described how tensor product algebra of a rank three tensor provides a framework for representing recursive structure in particular for language. However, the background work done with neural networks and tensors does not offer the tool necessary to resolve the problem faced here which is to add order to LSA.

There is no reason here to go any higher in rank than 2 because the background information LSA provides is limiting in that regard. One may then start with a mathematical description that will hopefully make tensors comprehensible. In a scalar field, a single number describes a point, while in an n-dimensional vector field, n-numbers are needed to describe a point. In a tensor field n-squared numbers are used to describe a point or n-cubed numbers, etc.

In other words, our tensor representations would be in the form of matrices because they add only one rank to vectors. This rank is utilized to describe word order in the sentence, therefore we have the words below described in 6 dimensional vector space in table 6.5. The word FLAG for example is represented by the list of digits shown in bold.

| | | Dimension 1 | Dimension 2 | Dimension 3 | Dimension 4 | Dimension 5 | Dimension 6 |
|---|---|---|---|---|---|---|---|
| Vector 1 | THE | 0.99 | 0.78 | 1.00 | 0.80 | 0.96 | 0.85 |
| Vector 2 | FLAG | **0.29** | **0.17** | **0.31** | **0.27** | **0.32** | **0.19** |
| Vector 3 | FLIES | 0.26 | 0.23 | 0.24 | 0.24 | 0.24 | 0.23 |
| Vector 4 | HIGH | 0.45 | 0.33 | 0.46 | 0.33 | 0.45 | 0.40 |

Table 6.5: Four words are shown, each described in 6-dimensional vector space

If the four words above are utilised as forming a sentence and we wish to retain the information of their order within the sentence, then the representation shown in table 6.6 is necessary.

| Tensor 1: THE FLAG FLIES HIGH | | | | | |
|---|---|---|---|---|---|
| 0.99 | 0.78 | 1.00 | 0.80 | 0.96 | 0.85 |
| 0.29 | 0.17 | 0.31 | 0.27 | 0.32 | 0.19 |
| 0.26 | 0.23 | 0.24 | 0.24 | 0.24 | 0.23 |
| 0.45 | 0.33 | 0.46 | 0.33 | 0.45 | 0.40 |

Table 6.6: The matrix as a whole describes the sentence maintaining the order of the words

In this representation, there is no division between the rows or columns so none of them can be isolated from the group. Additionally the representations for each word in the sentence are shown in the order in which it appeared, so the resulting matrix retains the information of where each word appeared in the sentence. It is this exact information that is lacking in LSA so if we refer to the prior example concerning Mark and the tiger, we find that if the word "*Mark*" appears first then the representation will have a different matrix from what it will have if the word "*tiger*" appears first because the order of the numbers in the rows is different.

In order to include tensors into a neural network model to form a novel architecture, one must have a solid background in tensor mathematics as well as some background on competitive architectures. Appendix C should provide a reasonable background in the mathematics involved in this task. It is worth mentioning that the functions mentioned reflect the specific requirements of this work, so the formulae although correct are not in their most general format in order to clarify them to the reader. Appendix D offers background information about how competitive architectures work to make it possible to focus here on the novel architecture.

## 6.4 A Tensor Based Competitive Architecture

The idea stems from the architecture of a self-organizing neural network algorithm offered by Tuevo Kohonen (1995) that allows inputs to organize themselves into different classifications in high dimensional space based on a competition of how close they are to each other. The complete mathematical basis for the way competitive architectures work is presented in Appendix D.

A central issue here is that once the network is settled after running many times, the inputs can be presented again to obtain an output representation. Let us assume that figure 6.1 shows the final positions of the weights. The three weights isolate three regions in the vector space using the weights w1, w2 and w3.

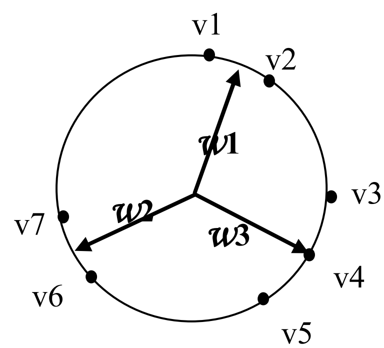



Figure 6.1: Weights isolating vectors in separate parts of the space

If we then present t the network vector v4, then its numerical representation is multiplied by the final weight values and the result is a string of values in terms of the weights. Since v4 and w3 have the same position then the resulting string will give v4's exact position.

However, if we do the same for v3 then the result will approximate v3's classification position at a value between it and the closest weight. Classification functions usually use a hard limit function to indicate if a vector belongs to that group or not. It has two values of either 1 for the vectors that belong there or 0 for those that do not and through this any approximations made are neglected at this stage. If, however, the function used is altered such that it is a continuous one, then the result would reflect that V3 did not end up as an exact match to a weight which in our case is a very important bit of information. This will be discussed in more detail in section 6.4.3.

The study involved several stages of analysis that took place. The first was to collect and assemble co-occurrence vectors to describe all the words that may appear in selection task rules in terms of each other. This selection task vocabulary was then used as a basic input source for the neural network and required necessary formatting into tensor matrix format. The third stage was to run the network to obtain classification results. The fourth was to compare results with LSA's results which were already reported in section 6.2.1.

### 6.4.1 Stage 1: Selection Task Vocabulary

Although the online site offered for LSA represents a window of opportunity for researchers, it does not allow one to extract the numerical representation for each word in a 300 dimensional vector form.

Consequently, it was necessary to manually obtain such a representation of words. Each of the words in the rules in Appendix A was counted to give a total of 218. A table was then formed of the co-occurrence similarity measure between that word and the other 217 words in the list while taking into account that its similarity with itself is 1. This was all done manually to gradually build a table of size 218x218 where each word can be represented in terms of all other words that appeared in the rules.

### 6.4.2 Stage 2: Formulation as Tensors

This second step involves building matrices to represent each antecedent and each consequent. First all rules were processed to remove the words "*if*", "*then*" and "*only if*" and to extract the antecedent and consequent clauses. Then for each word, the table described in section 6.4.1. was consulted to retrieve the word's vector representation.

These vector representations were arranged in such an order into the formed matrix to reflect the order of the words. Zeros were then added as fillers to ensure that all formed matrices have a size of 21x218 because this is the length of the longest clause in the study.

At the end of this stage, each antecedent and consequent had a matrix representation formed while respecting word order within the clauses which made them a total of 74 inputs.

### 6.4.3 Stage 3: Design of the Tensor Based Neural Network

Several parameters had to be set before starting the program including the function that will be used in the network, the number of weights and the number of repetitions of the program.

In classification type systems, the function that is frequently used is a hard limit function. What it does is compare the result to 0.5, so if the results are above it, the response is a 1 and if it is below it then the result is a 0. This implies that all vectors that appear within a particular region around the weight will get a classification factor that reflects they are of the same group because they will all get a 1 as a result while others will get a 0. Through this technique, the space is broken into zones and each is a classification to which some of the rules belong.

This type of classification is not desired here because it ignores part of the response which is up to 0.49 because of the rounding. Consequently, this function was omitted and the network was allowed to give the actual results obtained. Only a learning rate was utilised to make sure the network proceeds at small steps towards its target and this was set at 0.3.

The number of weights usually determines the number of classification regions. We have 37 rules, each of which is broken into two parts making the total number of inputs 74. This particular selection is well known to come through trial and error rather than by following any particular formula. The goal in this work is to have sufficient weights to check if rule antecedents and consequents are classified. At the same time, having too many weights would fail to classify because each weight may end up representing one input. Therefore the number of weights was set at 10.

The third parameter is to specify the number of times all the inputs are to be presented to the network. If the network's aim is to attain a particular target, then the network could be allowed to repeat until all distances between weights and inputs are below a pre-specified quantity. This, however, causes a problem if the specified quantity is not a practical one, as it will result in keeping the network in an infinite loop. This is especially the case if the input points are on both sides of one weight and at a distance larger than the one considered acceptable. In this case, the first input will bring the weight closer towards it and the second will change the same weight in the reverse direction to the first because it pulls it towards itself. The process will be endlessly repeated causing an infinite loop.

Since it is much more complicated to track these problems in the tensor world due to the high dimensionality involved, the total number of repetitions was fixed through trial and error to 19 total presentations. This causes all 74 inputs to be presented 19 times and each time is compared to 10 weights. Additionally, each input is represented by a

12x218 matrix, making the total number of comparisons involved 36,780,960. This should give the reader an idea of the processing power and high computational load of this network.

Now that all the basic parameters have been set, we can go on to review the different parts of the network. In the presentation step, the input tensor is presented to the network. This is followed by a competitive step where each weight tensor is compared to the input tensor and the closest weight is selected. Following this, the closest weight tensor is adjusted to bring it closer to the input such that the adjustment is restricted to a small distance. The idea is that with the repetitions, it will arrive at the optimal location.

Then, the second input is presented to the network and so on until all have been presented. This is called an epoch and is repeated 19 times as described above. Once the whole process is complete, the full list of inputs is presented one last time to the network except this time the results are reported rather than used to adjust weights.

However, these results are not interpretable as such because of the high dimensionality of the data. Consequently, another short program was written in order to find a similarity measure between the different rule parts. The output of this program is a table of size 74x74 to show the similarities between the rule clauses.

### 6.4.4 Stage 4: Results of Neural Network Compared to LSA

From the resulting matrix the similarity measure between the antecedent and consequent of each rule is reported in table 6.7, while compared to the corresponding figure from LSA.

| Deontic Rules | Premise Proximity Neural Network | No of words in antecedent – No of words in Consequent | Descriptive Rules | Premise Proximity Neural Network | No of words in antecedent – No of words in Consequent |
|---|---|---|---|---|---|
| **R2** | 23.89 | 7-10 | **R1** | 48.45 | 4-6 |
| **R4** | 23.89 | 7-10 | **R3** | 48.45 | 4-6 |
| **R5** | 105.78 | 6-5 | **R7** | 32.82 | 8-9 |
| **R6** | 105.78 | 6-5 | **R9** | 5.58 | 4-4 |
| **R13** | 23.89 | 7-10 | **R10** | 5.58 | 4-4 |
| **R14** | 49.2 | 7-11 | **R11** | 28.23 | 8-8 |
| **R15** | 29.75 | 8-5 | **R12** | 28.23 | 8-8 |
| **R16** | 29.75 | 8-5 | **R17** | 32.82 | 8-9 |
| **R21** | 48.52 | 7-10 | **R18** | 29.13 | 9-7 |
| **R22** | 39.74 | 8-5 | **R19** | 17.72 | 8-9 |
| **R38** | 81.24 | 6-9 | **R20** | 22.65 | 6-5 |
| **R42** | 81.24 | 6-9 | **R25** | 28.23 | 8-8 |
| **R46** | 81.24 | 6-9 | **R26** | 29.13 | 9-7 |
| **R56** | 81.7 | 5-8 | **R27** | 25.29 | 10-7 |
| **R57** | 99.92 | 4-6 | **R28** | 28.23 | 8-8 |
| **R58** | 81.7 | 5-8 | **R29** | 29.13 | 9-7 |
| **R59** | 99.92 | 4-6 | **R30** | 25.29 | 10-7 |
| **R60** | 51.8 | 6-12 | **R31** | 28.23 | 8-8 |
| **R61** | 55.98 | 10-7 | **R32** | 29.13 | 9-7 |
| **R62** | 51.8 | 6-12 | **R33** | 25.29 | 10-7 |
| **R63** | 55.98 | 10-7 | **R34** | 14.68 | 6-4 |
| **R66** | 70.61 | 5-6 | **R35** | 14.68 | 6-4 |
| **R67** | 70.61 | 5-6 | **R36** | 14.68 | 6-4 |
| **R68** | 24.88 | 6-9 | **R74** | 82.73 | 4-6 |
| **R69** | 24.88 | 6-9 | **R75** | 82.73 | 4-6 |
| **R70** | 37.93 | 9-12 | | | |
| **R71** | 37.93 | 9-12 | | | |
| **R72** | 26.83 | 6-9 | | | |
| **R73** | 26.83 | 6-9 | | | |
| **R76** | 48.59 | 8-10 | | | |
| **R77** | 48.59 | 8-10 | | | |
| **R78** | 19.5 | 7-9 | | | |
| **R79** | 19.5 | 7-9 | | | |
| | | | | | |
| **Mean** | **53.31** | | | **30.28** | |
| **St Dev.** | **27.49** | | | **18.76** | |

Table 6.7: The results of the neural network while considering antecedent-consequent rule lengths

The difference between the means for the NN is 43% with respect to Deontic rules which are much higher than LSA, which the reader may recall is 21%. There is no correlation between rule length and the results obtained from the neural network as the Pearson Product Moment Correlation coefficient is -0.24 which is extremely low. Consequently, one can go on to compare the graphs of the two as shown below in figure 6.2 and figure 6.3.

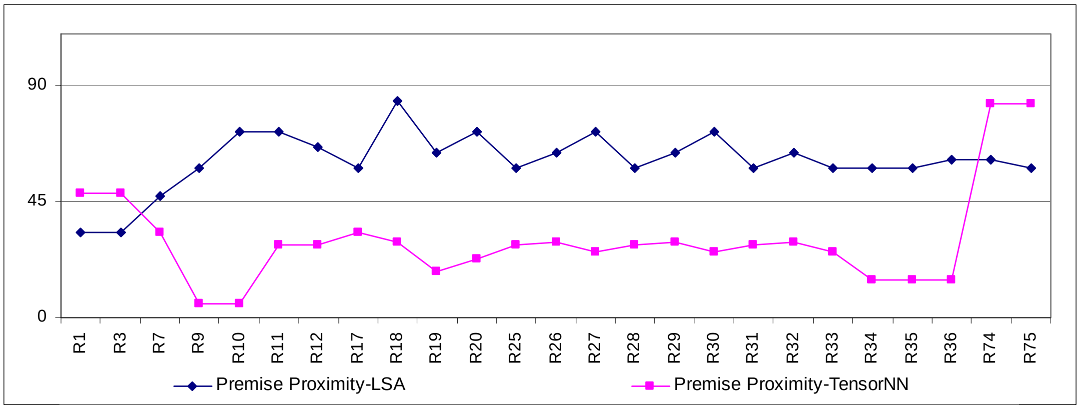



Figure 6.2: Descriptive proximity results from the NN compared to LSA

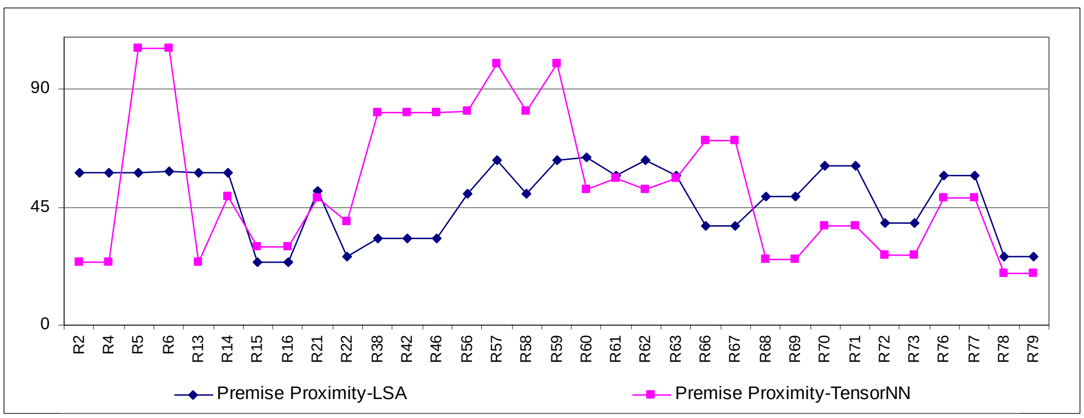



Figure 6.3: Deontic rule proximity results from the NN compared to LSA

A visibly clear result is that the neural network did not alter the descriptive representations as much as it did for the deontic representation. The graph shows a reflection around the 45% line that simply implies that in most cases the network resorted to multiplying the original with a fixed factor in order to obtain its results. However, the same could not be said for the deontic rules as they differ significantly from their original positions in a manner that is not organised.

### 6.4.5 Discussion

A central issue of concern here is the question of what exactly the network is isolating as a main difference between descriptive and deontic rules that causes the results reported. The results obtained for descriptive rules are almost identical to those obtained from LSA which forms the basic input resource to the network with the exception of a factor that reflected them around the 45% line.

This can only occur if the reader may recall when the weights coincide with the locations of the vectors or in this case tensors as explained in the introduction section 6.4. The network was run more than once and always kept repeating the same type of behaviour even though it was started from randomly generated initial weight positions. This indicates that the descriptive rule representations lie in positions at the centre of the isolated regions of classification which ends up as the position of the weight, while deontic rule representations occur around the weights, so they get approximated. An example of this is shown in figure 6.4

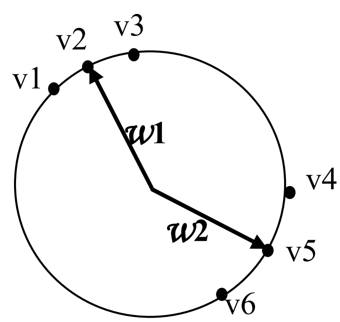



Figure 6.4: A representation in two dimensions of how weights approximate the regions.

Here each of w1 and w2 coincide with v2 and v5 which implies that results will reflect those vectors as they are while results will show approximate values for all of v1, v3, v4 and v6. The fact that weights will always coincide with descriptive rules, however, does isolate them into a group of their own that seems to partition the representation space.

## 6.5 Conclusion

This chapter has taken a look at the rule surface appearance which is the structure that holds the various referents to the values and positions that were analysed in the two prior chapters. The first result was that LSA is capable of correctly classifying rules as deontic versus descriptive with a reasonable degree of accuracy by showing that the two classes have different similarity means between the antecedent and consequent clauses. A pattern of words that frequently appear in the two categories started to emerge. Descriptive rules have a much higher tendency to contain "*general*" words that may appear in different contexts such as "*one, on, in, to, the*". Deontic rules, on the other hand, have a higher tendency of containing words such as "*cabin, employee, weekend*" which have a much lower use domain than do the general words.

Most of the general use words that seem to cause the distinction appeared as function words within the group of rules analysed here. Since LSA only analyses word frequencies and ignores their order, the role or significance of the difference may not be self evident just through the results exhibited, so a tensor-based NN was designed and utilised to classify the antecedents and the consequents by dividing the representational space using weights. The weight positions coincided with the descriptive rule antecedents and consequents almost without fail. This can only occur if these clauses represent the middle position of each class and therefore represent the optimal position in the centre of the group. Consequently, the high frequency function words that distinguish these rules are such that they occupy particular positions in the clauses they are part of that is otherwise occupied by open lexically classed words in their deontic counterparts.

The significance of this difference is difficult to determine for several reasons. The first is that the number of analysed rules is limited and consequently the phenomenon may be characteristic of just this domain of rules under analysis. The way to determine if this is the case, is by searching for a rule that has descriptive results and has enough open lexical class words to classify it as a deontic rule. The reader may be reminded that this did occur, in the case of the rule R1 which is "*If a letter is sealed, then it must carry a 20-cent stamp.*" When Johnson-Laird et al. (1972) tested this rule on British subjects, the percentage of correct responses was 81%. On the other hand, Griggs and Cox (1982) failed to observe facilitation by the same rule on American subjects. Golding (1981) later found that the postal rule produced better results when attempted by older British subjects who were familiar with a regulation that is no longer imposed by the British post office.

This implies that the only existing example of a rule that breaks the requirement in the form desired here does so because it can be affected by a context that affects subject behaviour and indeed the same goes for the rules that are classed as deontic and the network's results show them to be descriptive.

Of course it is hard to generalize rules that were not included in the study just by conjecture, but it should be safe to assume that given the current data, a basic form of descriptive rules does seem to exist such that it distinguishes it from deontic rules through the use of function words in particular positions.

An objection that may arise at this point is that LSA does not reflect the semantics of the words used in descriptive rules. Researchers including Smith and Witten (1993) identify function words as being closed class words which implies that although languages grow on a daily basis, function words are characterised by being both limited in number as well as used with high frequency. This implies that LSA is an ideal tool to isolate function words from open lexical words that carry more meaning just by identifying the frequency of their appearance which is what they conclude in their paper.

One of the main conclusions therefore made in this chapter, is that a distinction does seem to exist at the surface level of the rule structure that may in fact influence whether the rule is classed as deontic or descriptive. Whether it is real or just a phenomenon of the data analysed here remains to be seen as more experiments are run in the task. However, it does seem to indicate that the wording of the rules may affect the interpretations and to date no analysis has been carried out to identify the effects of the different lengths of rules for example, or different types of phrases used on subject reasoning.

# Chapter 7

# General Conclusion

Every show must eventually come to a close and with it, the memories of all the highlights come flashing in front of our eyes. The main aim of the work done in this thesis is to emphasise the importance of focusing on the interpretations subjects make in this and other reasoning tasks. The semantic implications of the questions' texts have the ability to influence the interpretations subjects make that in turn lead to a statistically significant difference in distribution among possible responses.

This thesis offers support to these claims through three main contributions. The first reveals novel interpretations by testing a rule format that describes a sequence of events. The second utilises information packaging as a semantic tool that is capable of identifying some of the distinctions that subjects impose onto the given information as well as identifying the effects of altering the letter/number hierarchy in the baseline descriptive task. Last but not least, part of this contribution reveals a serious misconception held by researchers to date by eliciting subject interpretations of logical negation to find that they too vary according to the assumptions made. The third is a neural network architecture that is capable of distinguishing deontic rules from descriptive rules based on co-occurrence information. In addition to that, descriptive rules are classified according to function words which add little to meaning as well appearing in a variety of different contexts.

This chapter will discuss these contributions in more detail and study their implications for research in this task and reasoning in general.

## 7.1 Novel Findings in The Selection Task

The first main finding is that subjects exhibit an unusual constant anaphor reading that is invited by implying a sequence in time through the use of different tenses. Subjects exhibited a suppositional reading of the conditional that is in a sense one of the primary semantic goals of the semantic construct.

> "*The use of the conditional* ***if*** *requires the listener to make suppositions, to entertain hypotheses or to consider once or future possible worlds*"
> (p. 29, Evans et al, 1993).

So the conveyor setting simply invited a particular interpretation and was unable to make the same invitation when the task became vague as was caused by the removal of the disjunction.

Information packaging offered a second perspective through which the semantics of the question could be contrasted without altering the propositional content of the task. Results indicated that subject attention rests on the cases that fit the rule by keeping them distinct from the cases that do not fit the rule. Consequently, altering this focus by emphasising a distinction between what the antecedent refers to and what the consequent refers to which is also done by emphasising the distinction between what lies on one side of the card versus the other side, results in directing subject behaviour towards a different distribution of responses. Cross-cutting the two possible lines of distinction reverts results to a distribution with a higher degree of confusion.

A few conditions that were originally designed as controls exhibited some of the necessary requirements for the question to invite interpretations leading to the competent responses. The hierarchy of the relationship between what the antecedent and consequent refer to was altered by making them both refer to letters and utilising colour to distinguish the face from the back of the cards. The referents therefore belong to the same domain as the distinction between the face and back is a trait that both can have.

The negation task is a highly insightful novel task that followed each of the conditions of the colour task. Results showed that current literature is erroneous in the assumptions of what subjects regard as negatives of P and Q, because responses varied across several choices that were explained through default logic. A central finding made here is that many subjects select P, Q while they interpret Q as the negative of P and P and the negative of Q. This implies that they are seeking to falsify by selecting a negative value in their selections which is something that will be picked up later in this chapter.

Perhaps, the division that lies between deontic and descriptive tasks is more or less profound than one might at first assume. Results of LSA shows that a distinction lies at the surface level represented by the distribution of types of words utilised in each of the two groups. The suspected culprit is an extensive use of "*function words*" in descriptive tasks, and this is further investigated through a tensor based neural network that exhibits the distribution of descriptive rules in a co-occurrence representation space as centre lines identifying the different domains. The distinction between the surface level of the two is therefore further fortified although there is no direct link indicating that this is indeed how subjects recognise the difference without any experimental evidence of the direct effect of function words.

## 7.2 What These Results Add to Existing Theories

It is essential to consider the informative extent of these findings in view of current theories. Results are not informative concerning the innateness or modularity of cheating detectors as suggested by Cosmides (1989) because the only review of deontic tasks that was done was through the neural network findings. However, they do raise questions whether a link can be established between function words and the scope of interpretations that subjects may arrive at or as to whether the two groups actually differ only in how clear each is in communicating what is desired of the subject. If this is the case, then that would question the existence of independent cheating detectors.

Relevance theory finds renewed support in the work presented here by the conditional that causes an increase in the logically competent answer and exhibits a necessity that the referents are closely related to each other by being members of the same domain. There is, however, no evidence that the reasoning that is required to arrive at the logically competent answers is any different in nature from that required to arrive at any other response category, as all seem highly sensitive to the distinctions emphasised in the task.

The theory of Mental Models is founded on the principle of truth (Johnson-Laird, 1999) where subjects only represent what is true. Results indicate that the distinctions assumed by researchers between what represents the negatives are not those made by most subjects who responded to the negation question. Johnson-Laird and Byrne (1991) also claim that subjects who do not turn the Not Q card do not represent it in their Mental Model. The assumption made is that subjects will identify the antecedent as P and the consequent as Q in order to place a distinction between them. However, a large proportion of subjects seem to regard Q as the negation of P and emphasise a distinction between what appears in the rule and what does not and this implies that they do not exhibit any problems with implicit negations. Additionally, the different conditions of the experiment described in chapter 5 produced variations in subject behaviour that remain unexplained by the theory of Mental Models.

The rational analysis model of the selection task associates probabilities of selection of each card with information gain. One of the main insights in the model is that it reflects a pragmatic view of reasoning such that subjects have a goal they attempt to achieve and in order to arrive at that goal they have to weigh the various possible choices. This is one of the basic assumptions of a theory of interpretation where subjects select the interpretation that seems most suited to their purposes and the selection is not always a certain one because they might alter it if they are confused one or more times.

The rational analysis model assumes that P and Q are independent as a basis for the model while in the results shown here a dependency does seem to exist with respect to contrast. Subjects place a distinction between the cases that fit the rule versus those that do not, and not between the referents of the antecedent and consequent. Another issue is raised by the results of the negation task. Not P is regarded as the information content of the value other than P rather than a symbolic representation of the negative of P. Consequently, subjects do indeed seem to wish to cover all possibilities of the antecedent by checking the value of P and what it is not which is well aligned with the information gain theory. This is not the case, though with deontic tasks as this particular interpretation of the consequent as a negative value of the antecedent is not likely to be made by subjects in this type of task which causes subjects to interpret the task in a format that is well aligned with what the experimenter intends.

Theories that assume a mental logic are also informed by these results because they indicate a strong effect of semantics when subjects interpret the task. In other words, a central problem seems to arise when the task given to subjects is regarded as requiring the interpretation of subjects and these do not necessarily conform to the interpretation assumed by the experimenter. Consequently, the selection of P, Not Q as the correct answers is based on an assumption of an interpretation that is not adopted by subjects according to the results presented here. Instead, results indicate that the interpretation step is extremely influential in guiding subject responses. In fact, there is no contradiction between the existence of a form of mental logic and the results of this thesis once the issue of interpretations is resolved.

## 7.3 What This Means to Reasoning

Perhaps a central issue that persevered throughout the years from the time when Wason ran his first selection task experiment to date is that most researchers continued to regard the task through a "looking glass" used by Wason. This looking glass dictated that the task has a "logical rule" and consequently its correct responses are those that can be deduced by classical logic. All responses other than P, Not Q were simply classified as wrong.

This restricted the work that followed into attempting to comprehend the cause of the widespread errors in subject responses. While in analogical reasoning researchers accepted the impreciseness of semantic implications, which causes subjects to interpret the stories to make deduction, this was not the case for selection task reasoning.

In order to arrive at the root of the problem and clarify the shortcomings of the looking glass, we may ponder what may have occurred if Wason had attempted the colour versions tested here, early on say around 1968. It is possible that he would only look at the correct/incorrect responses and simply say that they all fail in obtaining a sufficient correct ratio to warrant attention, but this is unlikely. The reason is that in his 1968 paper, he gave a full description of all responses given which implies that he is more likely to ponder subject distributions at least superficially.

If he did that, then it would be difficult for him to overlook the effects on subject behaviour of altering the semantics of the task. Clearly, the tasks here all fall under the same classical logical framework so responses would indicate that the differences warrant a more semantic based understanding of subject responses.

It is this understanding that presents the main thrust of the thesis. Simply, that it is impossible to comprehend the differences in the colour conditions, while assuming a classical logic perspective of the problem. The one approach that allows these results to be explained is through considering different subject interpretations that are affected by the semantic setting of the task given.

## 7.4 Future Work

This thesis opens many doors as it closes the door to my graduate study period. The most immediate question that bids attention is how directly linked is the formulation of the questions to subject responses. Can a framework of influencing effects be capable of "*directing*" thought towards particular predicted conclusions? This can only be achieved as a long term goal following many more specific intermediate goals.

These include running a detailed study of the effects of negative interpretations with a smaller number of conditions as well as running a version of the descriptive task to investigate the effect of having different tenses or having a causal relationship. These variations may guide work towards establishing a solid link from some of the more influential features of the question setting to the responses they invite.

The work will never be complete!!

*"We dance around in a ring and suppose, while the secret sits in the middle and knows" (Frost, 1971).*

# Appendix A

# Post-Hoc Study of Prior Experiments in the Selection Task

A careful review of a pool of the materials used in many of the papers referenced here was conducted. In order to establish a basis for comparison various variables had to be defined and given approximate values that usually ranged between two possibilities yes and no.

## A.1 Criteria of Evaluation:

These assignments were based on very strict criteria as the parameters and how they are assigned values is shown below:

1. Performance **Results** given as a percentage.
2. **Theme**: A rule is "thematic" if there exists an association in "*meaning*" and usually the original researchers reported this.
3. **Goal explicitly given**: If a person is told that their goal is to achieve something or enforce something.
4. **Social interaction** with a person: As per the requirements placed by Cosmides (1989).
5. **Assumed Background**: Decided based upon the report of researchers and how commonly known a theme is.

6. **Temporal Sequencing**: If any form of time is mentioned like days, weeks, etc or implication as when something occurs before another thing.

7. **Causal Link**: This can either be implied through the rule or through the context.

8. **Obverse possible**: If the rule or context implies that a possibility of cheating exists.

9. **Egoistic to subject**: Is there a "*you…are*" or role given to the subject performing the task

## A.2 List of Rules in the Review

The rules themselves were listed according to their RUNs. A run is an instance of attempting this experiment at a particular time and place. This makes it possible to compare results when the same rule is tested more than once by different researchers or in different settings. The list of runs analysed is as follows and the run number will appear in table A.1 containing the actual analysis results.

| **Materials** | **Cheng and Holyoak (1985)** |
|---|---|
| ***R1*** | Stamp problem: If a letter is sealed then it must carry a 20-cent stamp |
| ***R2*** | Cholera problem: if the form says entering on one side, then the other side includes cholera on its list of diseases |
| ***R3*** | Stamp problem: If a letter is sealed then it must carry a 20-cent stamp |
| ***R4*** | Cholera problem: if the form says entering on one side, then the other side includes cholera on its list of diseases |
| ***R5*** | If one is to take action 'A' then one must satisfy precondition 'P'. |
| ***R6*** | same as above when following abstract problem below |
| ***R7*** | If a card has an 'A' on one side, then it must have a '4' on the other side. |
| ***R8*** | when following permission case above |
| ***R9*** | If the tablecloth is brown then the wall is white. |

| | |
|---|---|
| ***R10*** | The tablecloth is brown, only if the wall is white. |
| ***R11*** | If a card has an 'A' on one side, then it has a '4' on the other side. |
| ***R12*** | A card has an 'A' on one side, only if it has a '4' on the other side. |
| ***R13*** | Cholera problem: if the form says entering on one side, then the other side includes cholera on its list of diseases |
| ***R14*** | Cholera problem: The form says entering on one side, only if then the other side includes cholera on its list of diseases |
| ***Materials*** | **Cheng et al. (1986)** |
| ***R15*** | If a customer is to drink an alcoholic beverage, then she is at least 18. |
| ***R16*** | A customer is to drink an alcoholic beverage, only if she is at least 18. |
| ***R17*** | If a card has an 'A' on one side, then it must have a '4' on the other side. |
| ***R18*** | If a bird has a purple spot underneath each wing, then it must build nests on the ground. |
| ***R19*** | If a washing label has silk on one side, then it has "dry clean only" on the other side. |
| ***R20*** | If two objects carry like electrical charges, then they will repel each other. |
| ***R21*** | If a passenger wishes to enter the country then he or she must have had an inoculation against cholera |
| ***R22*** | If a customer is to drink an alcoholic beverage, then she must be over 21. |
| ***R23*** | If a card has a circle on one side, then it has the word "red" on the other and conversely, if it has the world "red" on one side, then it has a circle on the other. |
| ***R24*** | If a turtle crosses a road, then the flag by the palace flies, and conversely if the flag by the palace flies, then a turtle crosses a road. |
| ***R25*** | If a card has an 'A' on one side, then it has a '4' on the other side |

| | |
|---|---|
| ***R26*** | If a bird has a purple spot underneath each wing, then it must build nests on the ground. |
| ***R27*** | If a bolt of cloth has any red threads in it, then it must be stamped with a triangle. |
| ***R28*** | If a card has an 'A' on one side, then it has a '4' on the other side |
| ***R29*** | If a bird has a purple spot underneath each wing, then it must build nests on the ground. |
| ***R30*** | If a bolt of cloth has any red threads in it, then it must be stamped with a triangle. |
| ***R31*** | If a card has an 'A' on one side, then it has a '4' on the other side |
| ***R32*** | If a bird has a purple spot underneath each wing, then it must build nests on the ground. |
| ***R33*** | If a bolt of cloth has any red threads in it, then it must be stamped with a triangle. |
| ***R34*** | If a house was built before 1979, then it has a fireplace. |
| ***R35*** | If a house was built before 1979, then it has a fireplace. |
| ***R36*** | If a house was built before 1979, then it has a fireplace. |
| ***R37*** | If a steel support is intended for the roof, then it must be rustproof. |
| ***R38*** | If any urithium miner gets lung cancer, then the company will pay the miner a sickness pension. |
| ***R39*** | If any of you wins an athletic award, then that person will have to treat the others to a round of drinks at Sam's |
| ***R40*** | If one works for the Armed Forces, then one must vote in the elections. |
| ***R41*** | If a steel support is intended for the roof, then it must be rustproof. |
| ***R42*** | If any urithium miner gets lung cancer, then the company will pay the miner a sickness pension. |

| | |
|---|---|
| ***R43*** | If any of you wins an athletic award, then that person will have to treat the others to a round of drinks at Sam's |
| ***R44*** | If one works for the Armed Forces, then one must vote in the elections. |
| ***R45*** | If a steel support is intended for the roof, then it must be rustproof. |
| ***R46*** | If any urithium miner gets lung cancer, then the company will pay the miner a sickness pension. |
| ***R47*** | If any of you wins an athletic award, then that person will have to treat the others to a round of drinks at Sam's |
| ***R48*** | If one works for the Armed Forces, then one must vote in the elections. |
| ***Materials*** | **Wason and Shapiro (1971)** |
| ***R49*** | If I go to Manchester, I travel by car. |
| ***R50*** | If I go to Manchester, I travel by train. |
| ***R51*** | If I go to Leeds, I travel by car. |
| ***R52*** | If I go to Leeds, I travel by train. |
| ***Materials*** | **Manketlow and Evans (1979)** |
| ***R53*** | If I eat haddock, then I drink gin (variations) |
| ***R54*** | If the letter is an N, then the number is a 3.(variations) |
| ***R55*** | If I go to Manchester, I travel by car. |
| ***Materials*** | **Gigerenzer and Hug (1992)** |
| ***R56*** | If a man eats cassava root, then he must have a tattoo on his face. |
| ***R57*** | If you eat duiker meat, then you have found an ostrich eggshell. |
| ***R58*** | If a man eats cassava root, then he must have a tattoo on his face. |
| ***R59*** | If you eat duiker meat, then you have found an ostrich eggshell. |

| | |
|---|---|
| ***R60*** | If someone stays overnight in the cabin, then that person must bring along a bundle of wood from the valley. |
| ***R61*** | If a student is to be assigned to Grover High School, then that student must live in Grover City. |
| ***R62*** | If someone stays overnight in the cabin, then that person must bring along a bundle of wood from the valley. |
| ***R63*** | If a student is to be assigned to Grover High School, then that student must live in Grover City. |
| ***R64*** | If a player wins a game then he will have to treat the others to a round of drinks at the club's restaurant. |
| ***R65*** | If a player wins a game then he will have to treat the others to a round of drinks at the club's restaurant. |
| ***R66*** | If a small-time drug dealer confesses, then he will have to be released. |
| ***R67*** | If a small-time drug dealer confesses, then he will have to be released. |
| ***R68*** | If an employee works on the weekend, then that person gets a day off during the week. |
| ***R69*** | If an employee works on the weekend, then that person gets a day off during the week. |
| ***R70*** | If a previous employee gets a pension from a firm, then that person must have worked for the firm at least ten years. |
| ***R71*** | If a previous employee gets a pension from a firm, then that person must have worked for the firm at least ten years. |
| ***R72*** | If a home owner gets a subsidy, then that person must have installed a modern heating system. |
| ***R73*** | If a home owner gets a subsidy, then that person must have installed a modern heating system. |
| ***R74*** | If an envelope is sealed, then it must have a 1-mark stamp. |

| | |
|---|---|
| ***R75*** | If an envelope is sealed, then it must have a 1-mark stamp. |
| ***R76*** | If a passenger is allowed to enter the country, then he or she must have had an inoculation against cholera. |
| ***R77*** | If a passenger is allowed to enter the country, then he or she must have had an inoculation against cholera. |
| ***R78*** | If a customer is drinking an alcoholic beverage, then he or she must be over 18 years old. |
| ***R79*** | If a customer is drinking an alcoholic beverage, then he or she must be over 18 years old. |

Table A.1: List of Rules Used in the Study

## A.3 Results of the Analysis

The result of the analysis is shown in table A.2 with the run number in the first column. Please note that most decisions made were according to the description given above and only set to a value of yes or no.

| | ***Results*** | ***Theme*** | ***goal Explicitly Given*** | ***social interaction with a person*** | ***Assumed Background*** | ***Temporal Sequencing*** | ***Causal Link*** | ***obverse possible*** | ***egoistic to subjects*** |
|---|---|---|---|---|---|---|---|---|---|
| **R1** | a86.5%[14] | yes | profit | yes | yes | yes | no | yes | yes |
| **R2** | a89.5% | yes | protect passengers from disease | yes | yes | yes | no | yes | yes |
| **R3** | a60% | yes | no | yes | no | no | no | yes | yes |
| **R4** | f57%t87.5% | yes | no | yes | yes | yes | no | yes | yes |

[14] a = About this amount, f=from, t= to this amount

| | ***Results*** | ***Theme*** | ***goal Explicitly Given*** | ***social interaction with a person*** | ***Assumed Background*** | ***Temporal Sequencing*** | ***Causal Link*** | ***obverse possible*** | ***egoistic to subjects*** |
|---|---|---|---|---|---|---|---|---|---|
| **R5** | 61% | no | no | yes | no | yes | yes | yes | yes |
| **R6** | 48% | no | no | yes | no | yes | no | yes | yes |
| **R7** | 19% | no | no | no | no | no | no | no | no |
| **R8** | 39% | no | no | no | no | no | no | no | no |
| **R9** | 17% | no | no | no | no | no | no | no | no |
| **R10** | 4% | no | no | no | no | no | no | no | no |
| **R11** | 17% | no | no | no | no | no | no | no | no |
| **R12** | 4% | no | no | no | no | no | no | no | no |
| **R13** | 67% | yes | no | yes | yes | no | no | yes | yes |
| **R14** | 56% | yes | no | yes | yes | no | no | yes | yes |
| **R15** | 67% | yes | no | yes | yes | yes | no | yes | yes |
| **R16** | 56% | yes | no | yes | yes | yes | no | yes | yes |
| **R17** | 19%+/- 6%[15] | no | | no | no | no | no | no | no |
| **R18** | 19%+/- 6% | %no | no | no | no | no | no | no | no |
| **R19** | 34%+/-6% | yes | no | no | no | no | no | no | no |
| **R20** | 34%+/-6% | yes | no | no | no | yes | Yes | no | no |
| **R21** | 66%+/-6% | yes | no | yes | yes | yes | yes | no | yes |
| **R22** | 66%+/-6% | yes | no | yes | yes | yes | yes | no | yes |

[15] With an error margin of + or – the percentage that follows.

| | Results | Theme | goal Explicitly Given | social interaction with a person | Assumed Background | Temporal Sequencing | Causal Link | obverse possible | egoistic to subjects |
|---|---|---|---|---|---|---|---|---|---|
| **R23** | 20%+/- 9% | no | no | no | no | no | no | no | no |
| **R24** | 20%+/- 9% | no | no | no | no | yes | no | no | no |
| **R25** | 15% | no | no | no | no | no | no | no | no |
| **R26** | 15% | no | no | no | no | no | no | no | no |
| **R27** | 40% | no | no | no | no | yes | no | no | no |
| **R28** | 40% | no | no | no | yes | yes | no | no | no |
| **R29** | 40% | no | no | no | yes | yes | no | no | no |
| **R30** | 50% | no | no | no | yes | yes | no | no | no |
| **R31** | 42% | no | yes | no | yes | no | no | no | yes |
| **R32** | 42% | no | yes | no | yes | no | no | no | yes |
| **R33** | 73% | no | no | no | yes | yes | no | no | yes |
| **R34** | 40% | yes | no | no | no | yes | no | no | no |
| **R35** | 50% | yes | no | no | yes | yes | no | no | no |
| **R36** | 73% | yes | no | no | yes | yes | no | no | yes |
| **R37** | 64% | yes | no | no | no | no | no | no | no |
| **R38** | 64% | yes | no | yes | no | yes | no | yes | no |
| **R39** | 64% | yes | no | yes | no | yes | no | no | no |
| **R40** | 64% | yes | no | yes | no | yes | no | no | no |

| | ***Results*** | ***Theme*** | ***goal Explicitly Given*** | ***social interaction with a person*** | ***Assumed Background*** | ***Temporal Sequencing*** | ***Causal Link*** | ***obverse possible*** | ***egoistic to subjects*** |
|---|---|---|---|---|---|---|---|---|---|
| **R41** | 57% | yes | no | no | yes | no | no | no | no |
| **R42** | 57% | yes | no | yes | yes | yes | no | yes | no |
| **R43** | 57% | yes | no | yes | yes | yes | no | no | no |
| **R44** | 57% | yes | no | yes | yes | yes | no | no | no |
| **R45** | 92% | yes | yes | no | yes | yes | no | no | yes |
| **R46** | 92% | yes | yes | yes | yes | yes | no | yes | yes |
| **R47** | 92% | yes | yes | yes | yes | yes | no | no | yes |
| **R48** | 92% | yes | yes | yes | yes | yes | no | no | yes |
| **R49** | 62.5% | yes | no | yes | yes | yes four trips on four days | no | no | yes |
| **R50** | 62.5% | yes | no | yes | yes | Yes four Trips on four days | no | no | yes |
| **R51** | 62.5% | yes | no | yes | yes | yes four trips on four days | no | no | yes |
| **R52** | 62.5% | yes | no | yes | yes | yes four trips on four days | no | no | yes |
| **R53** | 18.75% | yes | no | no | yes | yes four trips on four days | no | no | no |
| **R54** | 6.25% | yes | no | no | no | No | no | no | no |
| **R55** | 12.5% | yes | no | yes | no | yes four trips on four days | no | no | no |
| **R56** | 96% | yes | yes | yes | no | Yes | no | no | yes |

| | Results | Theme | goal Explicitly Given | social interaction with a person | Assumed Background | Temporal Sequencing | Causal Link | obverse possible | egoistic to subjects |
|---|---|---|---|---|---|---|---|---|---|
| **R57** | 91% | yes | yes | yes | no | Yes | no | no | yes |
| **R58** | 36% | yes | no | yes | no | Yes | no | no | yes |
| **R59** | 52% | yes | no | yes | no | Yes | no | no | yes |
| **R60** | 89% | yes | yes | yes | yes | Yes | no | yes | yes |
| **R61** | 77% | yes | yes | yes | yes | Yes | no | no | yes |
| **R62** | 53% | yes | no | yes | yes | Yes | no | yes | yes |
| **R63** | 46% | yes | no | yes | yes | Yes | no | no | yes |
| **R64** | 89% | yes | yes | yes | yes | Yes | no | yes | yes |
| **R65** | 41% | yes | no | yes | no | yes | no | no | yes |
| **R66** | 77% | yes | yes | yes | yes | yes | no | yes | yes |
| **R67** | 38% | yes | no | yes | no | yes | no | no | yes |
| **R68** | P&NotQ 75% | yes | employee | yes | yes | yes | no | yes | yes |
| **R69** | NotP&Q 61% | yes | employer | yes | yes | yes | no | yes | yes |
| **R70** | P&NotQ 70% | yes | employee | yes | yes | yes | no | yes | yes |
| **R71** | NotP&Q 64% | yes | employer | yes | yes | yes | no | yes | yes |
| **R72** | P&NotQ 81% | yes | employee | yes | yes | yes | no | yes | yes |
| **R73** | NotP&Q 59% | yes | employer | yes | yes | yes | no | yes | yes |
| **R74** | P&NotQ 80% | yes | postal worker | yes | yes | yes | no | yes | yes |

| | Results | Theme | goal Explicitly Given | social interaction with a person | Assumed Background | Temporal Sequencing | Causal Link | obverse possible | egoistic to subjects |
|---|---|---|---|---|---|---|---|---|---|
| **R75** | NotP&Q 28% | yes | sender | yes | yes | yes | no | yes | yes |
| **R76** | P&NotQ 85% | yes | immigration officer | yes | yes | yes | no | yes | yes |
| **R77** | NotP&Q 23% | yes | passenger that can cheat | yes | yes | yes | no | yes | yes |
| **R78** | P&NotQ 92% | yes | waitress | yes | yes | yes | no | yes | yes |
| **R79** | NotP&Q 11% | yes | customer | yes | yes | yes | no | yes | yes |

Table A.2: Results of the Data Analysis

In order to further clarify the data the **yes** was converted to a 1, and the **no** converted to a 0. Then the following groups were formed according to the results column and where that rule fits into a particular category. They were grouped as follows:

**Group A: Super-performer (percentage correct >= 80%):** This group included 15 runs and all fell into two main categories, either as social contracts with up to 95% accuracy rate or as permission or regulation schemas when a rationale is given. It may be worth mentioning that the social contract rules instructed subjects to look for cheaters.

**Group B: Moderate-performers (percentage correct < 80% and >= 56%):** This group included 29 runs with only one "*abstract*" case. It is the rule in Cheng and Holyoak (1985) "*If one is to take action 'A' then one must satisfy precondition 'P'*".

**Group C: Weak-performers (percentage correct < 56% and > 25%):** This group included 19 runs of which 8 are contingency training subjects and 8 are obligation training subjects of Cheng et al. (1986).

**Group D: Marginal-performers (percentage correct < 25%):** This group included 16 runs most of which had no clear relation between the two propositions.

When these results are sorted into four groups, figure A.1 results:

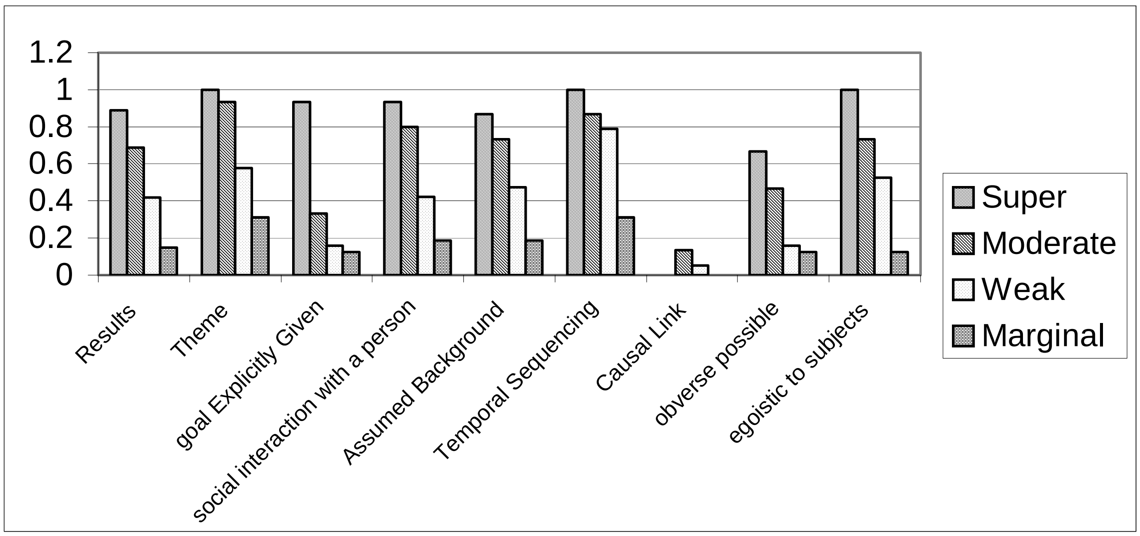



Figure A.1: Results of the Above Analysis Simplified into Four Groups

Some odd comparisons seems to emerge from this data, as we can clearly see that "*egoism*" and "*temporal distancing*" seem to correlate just as well with results as having "*thematic material*" or a "*social contract*". The odd thing here is that explicitly giving a goal does not seem to be doing that well in mimicking behaviour and only seems to be doing well in the case of the super performers. The Causal link test shows that this pool of data although diverse in the scope of results, does not show any influence of causal effects.

# Appendix B

# A Brief Description of LSA

Landauer and Dumais (1997) created a powerful high dimensional semantic model that bases its co-occurrent semantic information on the verbal contexts in which a word appears. This model uses contexts of up to 2000 words to describe the "*meaning*" of that word but it does not assign any importance to word order within that context.

The main basis on which LSA relies is that words, and any sets of words can be represented in an extremely high dimensional semantic space. What this means is that each word is represented through a vector which is composed of many numbers that describe how close it is to other words in that space. In a graphical sense, this implies that each word occupies a point in this multi-dimensional space that is described in terms of a number for each dimension. Figure B.1 shows how two vectors can be represented in a two dimensional space.

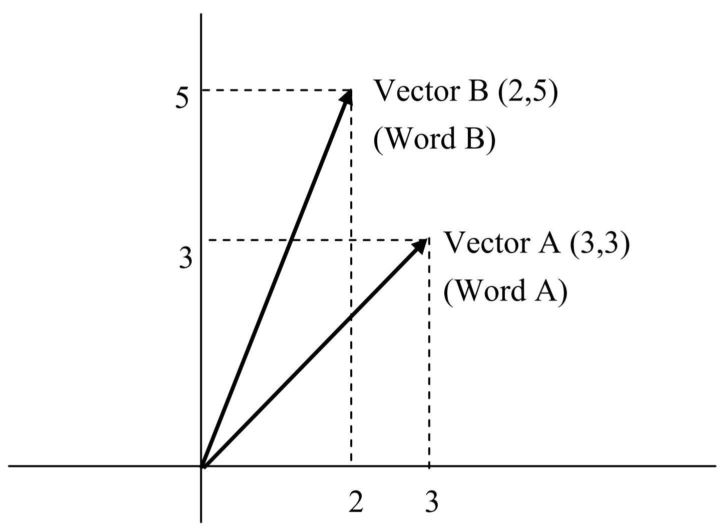



Figure B.1: Two vectors Represented in 2-Dimensional Space

Vector A (word A) therefore can be described by the values (3,3) in this world, while vector B (word B) can be described by the values (2,5). From here onwards cosine similarity is used to estimate distances between different words. This similarity measures the dot product of the two vectors divided by the product of the length of the vectors. The dot product of the two vectors above is computed as follows:

$A \cdot B = A_x B_x + A_y B_y = 3 * 2 + 3 * 5 = 6 + 15 = 21$

The product of the two lengths is computed as follows:

$|A| = SQRT\ (x^2 + y^2) = SQRT(\ 9 + 9) = 4.24$

$|B| = SQRT\ (x^2 + y^2) = SQRT(4 + 25) = 5.39$

Cosine similarity then becomes: 21 / (4.24 * 5.39) = 0.92

A high similarity measure implies that the two words often appear in similar contexts or, in other words, they both appear surrounded by the same words, so we expect the words "*elephant*" and "*zoo*" to have a high similarity measure, which is in fact 0.65, while oddly enough it is less likely to talk about "Tiger" and "zoo" in the same context with a similarity measure of 0.042.

This measure ensures that all results are between 0 and 1. However, this process can only take place after the model has been realized, with a process that is in no way trivial.

First of all, in LSA the world of words is far greater than two-dimensional ranging between 50 and 1500 dimensions (Landauer et al, 1998). The process starts through collecting 30,473 articles from Grolier's Academic American Encyclopaedia. From each article either the whole text was taken or up to 2,000 characters was taken with a mean length of 151 words. This was then used to fill a matrix that had 30,473 columns, each representing a piece of text. The matrix also had 60,768 rows, each representing a word that appeared in these texts. The fields were then filled in according to the function:

ln(1 + cell frequency) / entropy of the word over all contexts

The first part "ln" is a logarithmic function that uses a base of "e" which is 2.718 and is called the natural logarithm. One basic effect of using logarithms in general is to work with exponents since $\log_{10} 100 = 2$. This is then divided by the entropy of the word which is a measure of the "*importance*" of the word in the texts of that row and this we are told is measured with the function of sum of p log p (Landaur, et al. 1998). One possible interpretation of this is the following function:

$$\varepsilon_i = -\frac{1}{\log N} \sum_{j=1}^{N} f_{i,j} * \log f_{i,j} \qquad f_{i,j} = \frac{c_{i,j}}{m_i}$$

$c_{i,j}$ represents the frequency of the word in that particular text, while $m_j$ represents the frequency of the word across the texts. Applying the logarithmic function to this division raises the values to exponents and then applies a subtraction to them. An example in log base 10 is shown below:

$\log (100/1000) = \log 10^2 - \log 10^3 = 1 - 3 = -2$

Therefore the log of any division is in fact a method of subtraction of the two exponents of the numbers over the same base. When this number is then multiplied by the same fraction we get: -2*0.1 = -0.2 which is larger than the original fraction. This is then summed across the row and divided over the log of the number of entries in that row to get a more stable result also dependent on other entries in the same row.

Notice that if the two logs have equal exponents the result will be zero and when this zero is multiplied by the original fraction, which is 1, it will only end up as a zero. What this tells us is that word importance if it appears in only one text is reduced, while word importance rises if it appears in more than one text and in this way the "*importance*" of the word is incorporated into the cell entry.

A more complex process follows this initial stage and is explained through the example.

c1: Human machine interface for ABC computer applications

c2: A survey of user opinion of computer system response time

c3: The EPS user interface management system

c4: System and human system engineering testing of EPS

c5: Relation of user perceived response time to error measurement

m1: The generation of random, binary, ordered trees m2: The intersection graph of paths in trees

m3: Graph minors IV: Widths of trees and well-quasi-ordering

m4: Graph minors : A Survey

| | c1 | c2 | c3 | c4 | c5 | m1 | m2 | m3 | m4 |
|---|---|---|---|---|---|---|---|---|---|
| human | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| interface | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| computer | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| user | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| system | 0 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 |
| response | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| time | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| EPS | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| survey | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| trees | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| graph | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| minors | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

Table B.1: An Example of LSA/SVD offered by Landauer and Dumais (1997)

This table is then subject to a pre-processing stage as is described above. The full details of that stage, however, were never completely clarified and therefore it is difficult to fully understand the functions involved other than what is reported by authors. However, it was possible to reconstruct a table that results from applying the transformations as is shown in table B.2.

| | c1 | c2 | c3 | c4 | c5 | m1 | m2 | m3 | m4 |
|---|---|---|---|---|---|---|---|---|---|
| human | -0.1273 | -0.0329 | 0.0838 | -0.8647 | 0.5057 | 0.736 | -0.1257 | -0.5323 | 0.3331 |
| interface | -0.1969 | 0.2144 | -0.4472 | -0.7836 | 0.7813 | 0.3777 | -0.0521 | 0.3208 | -0.5056 |
| computer | -0.5071 | 0.1624 | -0.5677 | -1.0822 | -0.2932 | 0.2071 | -0.0062 | 0.0718 | 0.2251 |
| user | 0.2201 | 0.1628 | -0.2059 | -1.1774 | 0.0521 | -0.9157 | 0.5719 | 0.177 | -0.5116 |
| system | 0.785 | -0.5164 | 1.0166 | -1.8849 | 0.1125 | 0.4908 | 0.4706 | -0.0441 | -0.0632 |
| response | -0.0177 | 0.1575 | -0.2298 | -0.8558 | -0.3765 | -0.9009 | 0.4314 | -0.2995 | 0.0975 |
| time | -0.0177 | 0.1575 | -0.2298 | -0.8558 | -0.3765 | -0.9009 | 0.4314 | -0.2995 | 0.0975 |
| EPS | 0.5498 | -0.2359 | 0.5769 | -0.792 | 0.5904 | 0.3268 | 0.2094 | 0.1029 | -0.3949 |
| survey | 0.0895 | -0.2287 | -0.0572 | -0.8385 | -0.5037 | -0.5678 | -0.6123 | 0.4279 | 0.3379 |
| trees | 1.2826 | 0.0627 | -0.998 | 0.0873 | 0.0126 | 0.2465 | 0.1892 | 0.0394 | 0.4954 |
| graph | 1.1961 | 0.0759 | -0.6589 | -0.1699 | -0.1652 | -0.093 | -1.0088 | -0.1268 | 0.0857 |
| minors | 0.7494 | 0.5671 | -0.0793 | -0.1733 | 0.0074 | -0.2232 | -0.9284 | 0.199 | 0.3404 |

Table B.2: After applying the transformations as offered by Landauer and Dumais (1997)

*"We postulate that the power of the model comes from dimensionality reduction"(Landauer & Dumais, 1997)*

Following that, a Singular Vector Decomposition (SVD) is then utilized to break up the table into three. It turns the matrix shown above into three, such that when those are multiplied with each other the result is the original. SVD in a sense represents a matrix as a transformation from one vector space to another. The first and third matrices represent the vector spaces once in terms of the rows and the other in terms of the columns. The middle matrix represents a transformation matrix such that when it is reduced before multiplying the three together, this yields the best possible approximation to the original with respect to the selected number of inner dimensions. When the transformed data is broken up into its three components, it yields tables B.3, B.4 and B.5.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 0.22 | -0.11 | 0.29 | -0.41 | -0.11 | -0.34 | 0.52 | -0.06 | -0.41 |
| 0.2 | -0.07 | 0.14 | -0.55 | 0.28 | 0.5 | -0.07 | -0.01 | -0.11 |
| 0.24 | 0.04 | -0.16 | -0.59 | -0.11 | -0.25 | -0.3 | 0.06 | 0.49 |
| 0.4 | 0.06 | -0.34 | 0.1 | 0.33 | 0.38 | 0 | 0 | 0.01 |
| 0.64 | -0.17 | 0.36 | 0.33 | -0.16 | -0.21 | -0.17 | 0.03 | 0.27 |
| 0.27 | 0.11 | -0.43 | 0.07 | 0.08 | -0.17 | 0.28 | -0.02 | -0.05 |
| 0.27 | 0.11 | -0.43 | 0.07 | 0.08 | -0.17 | 0.28 | -0.02 | -0.05 |
| 0.3 | -0.14 | 0.33 | 0.19 | 0.11 | 0.27 | 0.03 | -0.02 | -0.17 |
| 0.21 | 0.27 | -0.18 | -0.03 | -0.54 | 0.08 | -0.47 | -0.04 | -0.58 |
| 0.01 | 0.49 | 0.23 | 0.03 | 0.59 | -0.39 | -0.29 | 0.25 | -0.23 |
| 0.04 | 0.62 | 0.22 | 0 | -0.07 | 0.11 | 0.16 | -0.68 | 0.23 |
| 0.03 | 0.45 | 0.14 | -0.01 | -0.3 | 0.28 | 0.34 | 0.68 | 0.18 |

Table B.3: Matrix W - representing the first component

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 3.34 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 2.54 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 2.35 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1.64 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1.5 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 1.31 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0.85 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.56 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.36 |

Table B.4: Matrix S - representing the transformation component

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 0.2 | 0.61 | 0.46 | 0.54 | 0.28 | 0 | 0.01 | 0.02 | 0.08 |
| -0.06 | 0.17 | -0.13 | -0.23 | 0.11 | 0.19 | 0.44 | 0.62 | 0.53 |
| 0.11 | -0.5 | 0.21 | 0.57 | -0.51 | 0.1 | 0.19 | 0.25 | 0.08 |
| -0.95 | -0.03 | 0.04 | 0.27 | 0.15 | 0.02 | 0.02 | 0.01 | -0.03 |
| 0.05 | -0.21 | 0.38 | -0.21 | 0.33 | 0.39 | 0.35 | 0.15 | -0.6 |
| -0.08 | -0.26 | 0.72 | -0.37 | 0.03 | -0.3 | -0.21 | 0 | 0.36 |
| 0.18 | -0.43 | -0.24 | 0.26 | 0.67 | -0.34 | -0.15 | 0.25 | 0.04 |
| -0.01 | 0.05 | 0.01 | -0.02 | -0.06 | 0.45 | -0.76 | 0.45 | -0.07 |
| -0.06 | 0.24 | 0.02 | -0.08 | -0.26 | -0.62 | 0.02 | 0.52 | -0.45 |

Table B.5: Matrix P - representing the last component

The matrix represented in table B.4 is described as the scaling matrix and it is critical to reproducing the original with a lower rank. A central feature of that matrix is that the values are getting smaller from the left to the right, which implies that when one ignores the columns to the right, then the factors ignored are those with a smaller effect on the total data. It is this feature that renders the singular vector decomposition with its ‘power’ by sorting the influence of factors in a decreasing fashion. Reduction therefore ends up in removing the factors that have the lowest influence on the overall data. Then if we wish to reduce the above matrix representation to two dimensions then redistributed, we do the following, reducing W to a two column matrix, and reducing S to a 2 by 2 matrix and reducing P to two columns. The result should be as is shown in table B.6.

| | c1 | c2 | c3 | c4 | c5 | m1 | m2 | m3 | m4 |
|---|---|---|---|---|---|---|---|---|---|
| human | -0.0235 | -0.0916 | 0.2205 | -0.6897 | 0.0954 | 0.0139 | 0.2524 | -0.0213 | -0.1111 |
| interface | 0.0251 | -0.0703 | 0.1624 | -0.6293 | 0.0707 | -0.0072 | 0.1967 | -0.0156 | -0.0828 |
| computer | 0.2223 | -0.0308 | 0.0374 | -0.7646 | 0.0187 | -0.0905 | 0.1006 | -0.0029 | -0.0237 |
| user | 0.3602 | -0.0543 | 0.0708 | -1.2738 | 0.0348 | -0.1465 | 0.1749 | -0.0057 | -0.0436 |
| system | 0.1641 | -0.2017 | 0.451 | -2.0178 | 0.1976 | -0.0587 | 0.5704 | -0.043 | -0.2319 |
| response | 0.3508 | -0.0066 | -0.0405 | -0.8651 | -0.0136 | -0.1448 | 0.0422 | 0.005 | 0.0129 |
| time | 0.3508 | -0.0066 | -0.0405 | -0.8651 | -0.0136 | -0.1448 | 0.0422 | 0.005 | 0.0129 |
| EPS | -0.0165 | -0.1206 | 0.288 | -0.9412 | 0.1248 | 0.0123 | 0.3333 | -0.0278 | -0.1455 |
| survey | 0.5586 | 0.0745 | -0.2657 | -0.6869 | -0.1089 | -0.2344 | -0.1686 | 0.0273 | 0.1225 |
| trees | 0.7659 | 0.2096 | -0.6186 | -0.0691 | -0.2597 | -0.3263 | -0.5292 | 0.0619 | 0.2967 |
| graph | 0.9873 | 0.2597 | -0.7727 | -0.1742 | -0.324 | -0.4201 | -0.6531 | 0.0774 | 0.3699 |
| minors | 0.7173 | 0.1883 | -0.5605 | -0.1295 | -0.235 | -0.3052 | -0.4735 | 0.0561 | 0.2683 |

Table B.6: Resulting matrix when the shaded areas in the paper are multiplied

Oddly enough, Landauer, Foltz and Laham (1998) gave the following matrix as a result:

| | c1 | c2 | c3 | c4 | c5 | m1 | m2 | m3 | m4 |
|---|---|---|---|---|---|---|---|---|---|
| human | 0.1637 | 0.4007 | 0.3743 | 0.4611 | 0.175 | -0.0531 | -0.1156 | -0.1585 | -0.0893 |
| interface | 0.1443 | 0.3773 | 0.3304 | 0.4016 | 0.1675 | -0.0338 | -0.0716 | -0.0969 | -0.0408 |
| computer | 0.1542 | 0.5062 | 0.3555 | 0.4095 | 0.2356 | 0.0193 | 0.0527 | 0.079 | 0.118 |
| user | 0.2581 | 0.8409 | 0.5947 | 0.6864 | 0.3908 | 0.029 | 0.0804 | 0.1212 | 0.1877 |
| system | 0.4534 | 1.2305 | 1.0394 | 1.2536 | 0.551 | -0.082 | -0.1686 | -0.225 | -0.0578 |
| response | 0.1636 | 0.5976 | 0.3785 | 0.4227 | 0.2832 | 0.0531 | 0.132 | 0.1913 | 0.2202 |
| time | 0.1636 | 0.5976 | 0.3785 | 0.4227 | 0.2832 | 0.0531 | 0.132 | 0.1913 | 0.2202 |
| EPS | 0.2217 | 0.5508 | 0.5071 | 0.6229 | 0.2414 | -0.0676 | -0.1464 | -0.2004 | -0.1083 |
| survey | 0.0991 | 0.5444 | 0.2335 | 0.221 | 0.2718 | 0.1303 | 0.3088 | 0.4392 | 0.4196 |
| trees | -0.068 | 0.232 | -0.1464 | -0.2682 | 0.1463 | 0.2365 | 0.548 | 0.7723 | 0.6623 |
| graph | -0.0678 | 0.3492 | -0.1433 | -0.2901 | 0.2106 | 0.2992 | 0.6942 | 0.979 | 0.8453 |
| minors | -0.0485 | 0.2554 | -0.1025 | -0.2088 | 0.1538 | 0.2172 | 0.5039 | 0.7107 | 0.6138 |

Table B.7: Resulting data when the two first rows in matrix P are selected and multiplied

These results in fact, emerge when one takes the first two rows of the last matrix instead of the first two columns as is highlighted in the paper. Additionally, in classical Singular Vector Decomposition, the multiplication is as follows: X = W*S*P' indicating that the last matrix is transposed or columns turned into rows and vice versa in the multiplication process. The results given in the paper arise from taking the first two rows and not applying any transpose to the data in the last matrix. However, since the power of the model has been demonstrated time and again throughout various publications, this may just be an error in reporting the example.

Therefore, this method of obtaining an approximation of the data is claimed to remove enough noise from the data to attribute proper descriptors of word meaning through the values that relate that word to its adjacent words. The ideal target dimensionality is also given as 300 dimensions.

Altering a single value in the original matrix, results in a change of a large group of cells in the resulting matrix. Results show a high degree of correlation between the predictions made by the model and human behavior. It also seems to have the ability of extracting a vector representation that is capable of assessing the "*semantic*" distances between words in a contextual space. The semantic space LSA uses to represent vectors, is a "*world of words*" where each word has a location based on the distance its meaning is from other words. This distance is estimated without any regard to where the word appeared in the sentence or as Landauer and Dumais (1997) described, it treats contexts as "*bags of words*". So it should not be surprising to find that the LSA model detects the following two sentences identical in meaning.

***Mark killed the tiger vs The tiger killed Mark***

It should be clearly evident that the meaning understood by the reader of the first sentence in the case of Mark killing the Tiger is extremely different from that the Tiger killed Mark as the death of a human is never equated with the death of a vicious animal.

This should be enough to emphasize the importance of word order in meaning, although the roles are not particularly clear at this stage. This importance was identified by Wiemer-Hastings (2000) to further investigate the effects of the neglected syntactical information in a sentence and to attempt to incorporate it into the LSA framework. He separated the sentences into atomic clauses or propositions and then segmented them by hand to break them into strings composed of subject noun phrase, verb and object noun phrase. Antecedents were used to resolve pronouns and conjunctions were dealt with by distributing the arguments. Then he attempted to evaluate the similarity of this presentation using a variety of measures. Results showed that the best approach to combine the similarities of the sentence parts is non-linear and even that was not as close to human judgment as LSA. Wiemer-Hastings and Zipitria (2001) then went on to a further test, incorporating syntax through two methods. The first was to tag the words used for the training corpus at 100, 200, 300 and 400 dimensions but this did not produce any favorable results. Then they tried a structured LSA or SLSA where they broke up sentences into parts as was shown above and used that as training material to find results that correlate slightly better than LSA, which does not pay any attention to sentence structure.

To sum up, although LSA offers itself as a computationally powerful model that extracts its knowledge from a large encyclopaedia, it does not take word order into consideration.

# Appendix C

# A Brief Review of Tensor Mathematics

Unfortunately, the name *tensor* seems to be constantly linked with Riemann geometry and the theory of Relativity. This strong relationship seems to invoke an impression that whatever is linked to this type of mathematics must be "*incomprehensible*". The truth of the matter lies in the fact that any topic that is regarded from the outside, seems to be "*incomprehensible*", so the best way to understand tensors is to visit some of the tensor representations that we have in our world. We can begin by visiting a beautiful Swiss Alpine mountain, for example, that can be seen from the top of another mountain. What one can see is represented in three dimensions, length, breadth and breathtaking height. The view itself allows one to see the valleys and the trees of various heights and colours. If this were captured in a matrix, then would it suffice to describe each point in that space just in terms of its height? Obviously not, the whole setting has to be retained keeping the length, breadth as well as height. This representation retains the height of each level of land from the top of the mountain relative to each other point so plateaus can be easily identified. This, is a tensor of the lowest rank!

We shall not endeavour here to go any higher in rank than that, as there does not seem to be any need to do so, but it is worth mentioning that the sky is the limit for anyone who wishes to incorporate more information into a description that this physical world offers. It is from this flexibility that tensors gained their widely renowned reputation of being complex.

Now in order to start with a mathematical description of a tensor that makes it more comprehensible, it could be compared to its peers in the math world. In a scalar field, a single number describes a point, while in an n-dimensional vector field, n-numbers are needed to describe a point. In a tensor field n-squared numbers are used to describe a point or n-cubed numbers, etc.

In other words, our tensor representations would be in the form of matrices because they add only one rank to vectors. This rank is utilized to describe word order in the sentence, therefore if we have the words below described in 6 dimensional vector space in table C.1 then a tensor representation of that sentence can be seen in table C.2 where the whole sentence is described by a matrix.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Vector 1 | THE | 0.99 | 0.78 | 1.00 | 0.80 | 0.96 | 0.85 |
| Vector 2 | FLAG | 0.29 | 0.17 | 0.31 | 0.27 | 0.32 | 0.19 |
| Vector 3 | FLIES | 0.26 | 0.23 | 0.24 | 0.24 | 0.24 | 0.23 |
| Vector 4 | HIGH | 0.45 | 0.33 | 0.46 | 0.33 | 0.45 | 0.40 |

Table C.1: Four words are shown, each described in 6-dimensional vector space

| Tensor 1: THE FLAG FLIES HIGH | | | | | |
|---|---|---|---|---|---|
| 0.99 | 0.78 | 1.00 | 0.80 | 0.96 | 0.85 |
| 0.29 | 0.17 | 0.31 | 0.27 | 0.32 | 0.19 |
| 0.26 | 0.23 | 0.24 | 0.24 | 0.24 | 0.23 |
| 0.45 | 0.33 | 0.46 | 0.33 | 0.45 | 0.40 |

Table C.2: The matrix as a whole describes the sentence maintaining the order of the words

## C.1 Tensor Subtraction

The origins of the numbers will be explained in the following section. For now, one can go a little deeper in some tensor mathematics. Tensors are basically constructs that obey certain transformations, allowing them to interact in a world that may seem strange to some. What would it mean to subtract one tensor from another? If our tensor is represented by a mountain, then it should be similar to subtracting a mountain from another, taking into account the locations of hills and valleys. One could imagine two hollow mountains and the process of making one semi-invisible and passing it through the matter of the other in order to identify a representation for the "*difference*" between the two tensors. In a vector world, this difference would be a third vector, so it should not be surprising that in a tensor world, the difference would be a tensor.

An equation for this follows:

$$\underline{\underline{T}} - \underline{\underline{U}} = \sum_{i=1}^{12}\sum_{j=1}^{218} T_{ij} - U_{ji}$$

The subscripts shown are those related to the tensor based neural network that will be explained shortly. Basically, what the summations are doing is to deduct each element belonging to U from its corresponding element belonging to T. Therefore from the point of view of calculations it is quite straightforward to carry out that formula computationally while it is at the same time powerful in that it is capable of subtracting two tensor based representations from each other. It should be clear by now, that the imagined view of a mountain is in fact more complex than one would think. Two of the dimensions are in fact a "*flat*" representation of a multi-dimensional vector world, which one assumes to be "*flat*" as a base to the mountain in order to add height to that mountain. Consequently, one may say that the base of the mountain is a world as complex as the one described by LSA while the added dimension is orthogonal to that world.

The tensor functions that are necessary for the model also include two main operations: estimating the magnitude of a tensor and calculating the dot product of a tensor.

## C.2 Tensor Magnitude

Yet again the mathematics of estimating the magnitude of a tensor is far simpler to understand than the power of the operation. The formula is shown below:

$$\left|\underline{\underline{T}}\right| = \sqrt{\frac{1}{2}\sum_{i=1}^{12}\sum_{j=1}^{218} T_{ij}^{2}}$$

Let us now contrast this formula with the one that measures the magnitude of a vector which is as follows:

$$\left|\underline{v}\right| = \sqrt{\sum_{i=1}^{25} v_i^2}$$

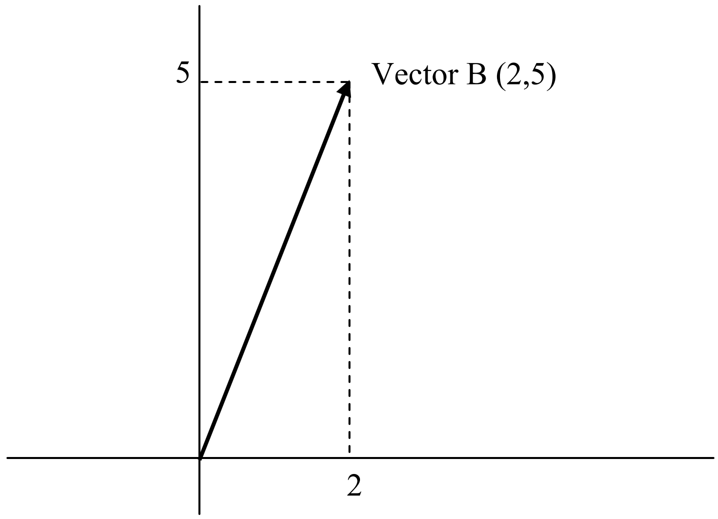



Figure C.1: A Simple Vector Representation

The magnitude of the two dimensional vector shown above can be calculated by squaring the components and getting the square root of the total. In other words the magnitude of Vector B equals

$$\sqrt{(5x5 + 2x2)} = \sqrt{29} = 5.385$$

When it comes to a tensor though, things become more complicated because of the higher level of dimensionality. Therefore, the calculations have to take into account the sum of squares on a per row basis. Once that sum is calculated, one also has to account for these values that are part of a larger construct, so they in turn have to be summed and only then can the magnitude of the tensor be calculated.

## C.3 Tensor Dot Product

The third necessary mathematical formula for the design of a tensor based competitive neural network is that of a dot product. Here two tensors are multiplied by each other to result in a scalar value. An example of this formula is as follows:

$$\underline{\underline{T}} : \underline{\underline{U}} = \sum_{i=1}^{12} \sum_{j=1}^{218} T_{ij} U_{ji}$$

Mathematically, this is easy to calculate but explaining it is a very different proposition. When it comes to vectors, the dot product of two vectors represents a good measure of orthogonality. If we get the dot product of two vectors with indices of (1,0) and (0,1) then it will be a zero because one vector lies on the positive side of the x axis while the other on the positive side of the y axis. This behaviour is repeated no matter where the vectors are in the representation so long as they are perfectly orthogonal to each other. In other words, even when we are talking about a complex tensor world, then orthogonality is detected. If a zero does not result, then the resulting number reflects the magnitude of the projection of one vector along the other. I will not attempt to explain this in terms of tensors for reasons that should be clear with regards to the complexity in hand. The point, however, should be clear, that this measure is extremely sensitive to direction.

Although discussing tensor mathematics is a great deal of fun, we should not over indulge ourselves in this wonderful world that exceeds imagination and go on to the competitive based architecture where these three operations will be of critical importance.

# Appendix D

# A Brief Description of Competitive Architectures

## D.1 How Competitive Architectures Work

In a simple competitive architecture, input patterns are presented to the network during a training phase. The weight vector is then adjusted to be able to differentiate between the vectors that are highly dissimilar and to group the ones that are highly similar.

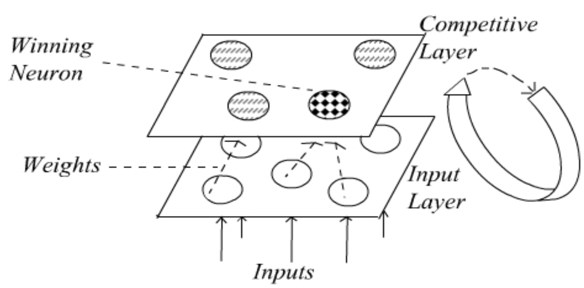


Figure D.1: A Competition Based Architecture

The process is clarified in figure D.1; as each input is presented to the network through its input neurons, a selected weight wins if it is the closest weight vector to that input. This weight is then adjusted in order to include the pattern into a classification that covers that area of space that the input is closest to. In a strange sense, this is a measure of how close vectors are to each other as is shown below in a simplification of how weights can classify two dimensional vectors.

v1

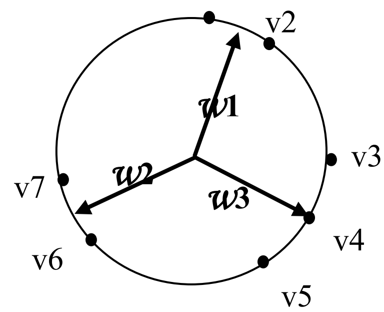



Figure D.2: Weights isolating vectors in separate parts of the space

The weights attempt to isolate three regions in the vector space such that each region classifies a group of vectors that are sufficiently similar to each other. The distance between vectors and weights is usually computed using a formula as follows:

$$\underline{v} - \underline{w} = \sum_{i=1}^{2616} v_i - w_i$$

Following the deductions the magnitude of each resulting difference is calculated in order to judge which weight wins such that the winner is the weight with the lowest difference. Note that the vector magnitude formula is also necessary to decide on the selection of the winning weight vector. Once it is selected, the weight vector is adjusted to be as close as possible to the input vector by adding an increment of a "*step*" further towards the input pattern. The process continues by presenting the whole data set over and over. Each presentation of the whole data set is called an "*epoch*". Following a sufficient number of epochs the resulting classification can be tested in order to identify the groupings. Additionally, through using the dot product formula on the whole data set, a similarity measure can be extracted from the resulting data showing the "*distances*" between the different vectors following classification.

An example of this process can be exhibited through the classification process of the following six vectors (Hagan, Demuth, Beale, 1996).

$$\mathbf{p}_1 = \begin{bmatrix} -0.1961 \\ 0.9806 \end{bmatrix} \quad \mathbf{p}_2 = \begin{bmatrix} 0.1961 \\ 0.9806 \end{bmatrix} \quad \mathbf{p}_3 = \begin{bmatrix} 0.9806 \\ 0.1961 \end{bmatrix}$$

$$\mathbf{p}_4 = \begin{bmatrix} 0.9806 \\ -0.1961 \end{bmatrix} \quad \mathbf{p}_5 = \begin{bmatrix} -0.5812 \\ -0.8137 \end{bmatrix} \quad \mathbf{p}_6 = \begin{bmatrix} -0.8137 \\ -0.5812 \end{bmatrix}$$

The competitive network offered to classify these has three neurons and therefore has three weight vectors that are initialised to the following values.

$${}_1\boldsymbol{w} = \begin{bmatrix} 0.7071 \\ -0.7071 \end{bmatrix} \quad {}_2\boldsymbol{w} = \begin{bmatrix} 0.7071 \\ 0.7071 \end{bmatrix} \quad {}_3\boldsymbol{w} = \begin{bmatrix} -1.0000 \\ 0.0000 \end{bmatrix}$$

Making the total weight vector $\mathbf{W} = \begin{bmatrix} {}_1\boldsymbol{w}^\mathrm{T} \\ {}_2\boldsymbol{w}^\mathrm{T} \\ {}_3\boldsymbol{w}^\mathrm{T} \end{bmatrix} = \begin{bmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \\ -1.0000 & 0.0000 \end{bmatrix}$

If we present the second input pattern to the network, it starts by estimating the distance between each of its weight vectors and that input pattern. Although there are several different ways of performing this comparison, the way shown here is the one followed in the model.

$$A = \begin{bmatrix} \sqrt{\Sigma\,({}_1\boldsymbol{w}_\mathrm{i} - \mathbf{p}_\mathrm{i})^2} \\ \sqrt{\Sigma\,({}_2\boldsymbol{w}_\mathrm{i} - \mathbf{p}_\mathrm{i})^2} \\ \sqrt{\Sigma\,({}_3\boldsymbol{w}_\mathrm{i} - \mathbf{p}_\mathrm{i})^2} \end{bmatrix} = \begin{bmatrix} \sqrt{(0.7071-0.1961)^2 + (-0.7071-0.1806)^2} \\ \sqrt{(0.7071-0.1961)^2 + (\,0.7071-0.1806)^2} \\ \sqrt{(-1.0000-0.1961)^2 + (\,0.0000-0.1806)^2} \end{bmatrix}$$

$$A = \begin{bmatrix} 1.7634 \\ 0.5796 \\ 1.5467 \end{bmatrix}$$

The closest weight is therefore weight 2 which is on the second row so its weight will be adjusted according to the following equation and here we assume the learning rate to be $\alpha$ = 0.5. This is larger than the learning factor used in the model to show a clearer change in the example. The model itself has a learning rate of 0.3.

$${}_2\boldsymbol{w}^{new} = {}_2\boldsymbol{w}^{old} + \alpha\,(p_2 - {}_2\boldsymbol{w}^{old})$$

$$= \begin{bmatrix} 0.7071 \\ 0.7071 \end{bmatrix} + 0.5 \left( \begin{bmatrix} 0.1961 \\ 0.9806 \end{bmatrix} - \begin{bmatrix} 0.7071 \\ 0.7071 \end{bmatrix} \right) = \begin{bmatrix} 0.4516 \\ 0.8439 \end{bmatrix}$$

Now if one compares the new weight, it is clearly closer to pattern two than its previous values were. If this process is repeated as all different patterns are presented repeatedly to the network, the weight swings like a pendulum until it arrives at a stable point at the centre of the region it covers. In this way, the different weights divide to cover different regions of the search space.

## D.2 A Tensor Based Competitive Architecture

Since the basic competition based architecture has been clearly outlined above, this section will concentrate on differences between that architecture and a tensor based one. The first clear difference is that of dimensionality, which we can see if we review our examples of mountains. In a vector based architecture the network compares vectors represented in multi-dimensional space and moves the weights in that number of dimensions. So on paper our arrows can move around to isolate different regions of classification as necessary. On the other hand, in a tensor based architecture, we add a dimension orthogonal to the paper drawing our imaginary mountains and then weights also have to be represented by "*mountains*" that separate these mountains into different classifications. Although it is difficult to imagine, this is what happens to a multi-dimensional vector space that is in a strange sense considered "*flat*" when the extra dimension is added through the tensor representation.

The main complications that arise after jumping from vectors to tensors is that the lack of visualization makes it more difficult to properly comprehend the full workings of that world. However, one must not forget that the weight values here are tensors and therefore complex constructs are required to break up the world into different "*zones*" that would contain the various classifications.

Therefore the tensor formulae described above go into this code in order to perform all the main operations necessary for the network. This requires a little bit of juggling during the design phase of this architecture a main reason being that the weight matrix W for example will become three dimensional rather than two dimensional as shown above. This in turn requires some decisions to be made as to the arrangement of the two dimensional matrices of weights with respect to their order in that weight matrix. Yet again, it takes another decision of how to make an effective multiplication considering that in the classical case each transpose of a weight is multiplied by the input pattern.

# Bibliography

Babcock, M.K. and Freyd, J.J. (1988). Perception of dynamic information in static handwritten forms. *American Journal of Psychology*, 101: 111-130.

Braine, M.D.S. (1978). On the Relationship between the natural logic of reasoning and standard logic. *Psychological Review*, 85: 1-21.

Cheng, P.W., Holyoak, K.J. Nisbett, R.E. and Oliver, L. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning, Cognitive Psychology, 18: 293-328.

Cheng, P.W. and Holyoak, K.J. (1985). Pragmatic Reasoning Schemas, *Cognitive Psychology*, 17: 391-416.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31: 187-276.

Follesdal, D. and Hilpinen, R. (1971) Deontic logic: An introduction. In R. Hilpinen, editor, Deontic Logic: Introductory and Systematic Readings, pages 1-35. Reidel, 1971.

Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W. and Harshman, R. A. (1991). Indexing by latent semantic analysis. *Journal of the American Society of Information Science*, 41(6): 391-407.

DesCartes, R. (1641). Meditations, DesCartes Meditations, translated by John Veitch (1901), A Trilingual HTML Edition, Manley, D. B. and Taylor, C. S. (Eds).

DesCartes, R. (1954). "Rules For The Direction Of The Mind." *Descartes Philosophical Writings*, pages 153-180. Translated by Elizabeth Anscombe and Peter Thomas Geach. Thomas Nelson and Sons Ltd. London, Evans, J.St.B.T. (1973). Matching bias in the selection task. *British Journal of Psychology*, 64: 391-397.

Evans, J.St.B.T. (1983). Linguistic determinants of bias in conditional reasoning. Quarterly Journal of Experimental Psychology, 35(A): 635-644.

Evans, J.St.B.T. (1993). The mental model theory of conditional reasoning: critical appraisal an revision, *Cognition*, 48: 1-20.

Evans, J.St.B.T. (1996) Deciding before you think: Relevance and reasoning in the selection task, *British Journal of Psychology*, 87: 223-240.

Evans, J.St.B.T. and Newstead, S.E. (1977). Language and reasoning: A study of temporal factors. *Cognition*, 5: 265-283.

Evans, J.St.B.T., Handley, S.J., and Buck, E. (1998). Ordering of information in conditional reasoning, *British Journal of Psychology*, 89(3): 383-404.

Evans, J.St.B.T., Newstead, S.E. and Byrne, R.M. (1993). Human Reasoning: The Psychology of Deduction, Lawrence Erlbaum Associates, Publishers, Hove, UK.

Fauconnier, G. (1994). Mental Spaces: Aspects of meaning construction in natural language, Cambridge University Press, Cambridge, UK.

Fillenbaum, S. (1978). How to do some things with if. In Cotton and Klatsky (eds.) *Semantic functions in Cognition*. Lawrence Erlbaum Associates.

Freyd, J.J. (1983a). Representing the dynamics of a static form. *Memory and Cognition*, 11: 342-346.

Freyd, J.J. (1983b). The mental representation of movement when static stimuli are viewed. *Perception and Psychophysics*, 33: 575-581.

Freyd, J.J. and R. A. Finke, (1984). Representational Momentum, *Journal of Experimental Psychology: Learning Memory & Cognition*, 10(1): 126-132.

Freyd, J.J. (1987). Dynamic Mental Representations, *Psychological Review*, 94(4): 427-438.

Freyd, J.J. (1993). Five hunches about perceptual processes and dynamic representations. In Meyer, D. and Kornblum, S. (Eds) *Attention & Performance XIV: Synergies in*

*Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience - A Silver Jubilee*, pages 99-119. Cambridge, MA: MIT Press.

Freyd, J.J., Pantzer, T.M. and Cheng, J.L. (1988). Representing statics as forces in equilibrium. *Journal of Experimental Psychology: General*, 117: 395-407.

Frost, R. (1971). The Road not Taken, Henry Holt and Company, New York.

Gebaur, G. and Laming, D (1997). Rational choices in Wason's selection task *Psychological Research*, 60: 284-293.

Gentner, D., Rattermann, M. J. and Forbus, K. (1993). The roles of similarity in transfer: separating retrievability and inferential soundness. Cognitive Psychology, 25: 524-575.

Gigerenzer, G. and Hug, K. (1992), Domain-specific reasoning: social contracts, cheating and perspective change. *Cognition*, 43: 127-171.

Gilhooly, K.J. and Falconer, W.A. (1974) Concrete and Abstract terms and relations in testing a rule, Quarterly Journal of Experimental Psychology, 26: 355-359.

Glenberg, A. M., and Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language,* 43(3): 379–401.

Golding, E. (1981) The effect of past experience on problem solving, *Annual Conference of the British Psychological Society*, Surrey University.

Griggs R.A. and Cox, J.R. (1982). The elusive thematic materials effect in Wason's selection task. *British Journal of Psychology*, 73: 407-420.

Hagan, M., H. Demuth, H. and Beale, M., (1996). Neural Network Design, PWS Publishing Company, Boston, Ma.

Handley, S.J. and Evans, J.St.B.T. (2000). Supposition and representation in human reasoning. *Thinking & Reasoning,* 6: 273-312.

Johnson-Laird, P., Legrenzi, P. and Legrenzi, S. (1972). Reasoning and a sense of reality. *British Journal of Psychology*, 63: 395-400.

Johnson-Laird, P.N. and Byrne, R.M.J. (1991). Deduction, Hove, UK: Lawrence Erlbaum Associates Ltd.

Johnson-Laird, P.N. and Byrne, R.M.J. (1993). Models and deductive rationality. In K.I. Manktelow and D.E. Over (Eds.), *Rationality: Psychological and Philosophical Perspectives*, London: Routledge.

Johnson-Laird, P.N. (1999). Deductive Reasoning. *Annual Reviews of Psychology*, Highwire Press, Stanford University.

Kirby, K.N. (1994). Probabilities and utilities of fictional outcomes in Wason's four-card selection task, Cognition, 51(1): 1-28.

Kohonen, T. (1995), Self-Organizing Maps, Springer, Berlin.

L. van der Torre and Y. Tan, (1998). The Temporal Analysis of Chisholm's paradox. In *Proceedings of 15th National Conference on Artificial Intelligence (AAAI'98)*, pages 650-655.

Landauer, T. K., and Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. P*sychological Review,* 104: 211-240.

Landauer, T.K., Foltz, P.W., and Laham, D., (1998). Introduction to Latent Semantic Analysis, Discourse Processes, 25: 259-284.

Mally, E. (1926). Grundgesetze des Sollens: Elemente der Logik des Willens. Graz: Leuschner und Lubensky, Universitäts-Buchhandlung, viii+85 pp. Reprinted in K. Wolf and P. Weingartner (Eds.) Ernst Mally, Logische Schriften: Großes Logikfragment, Grundgesetze des Sollens, pages 227-324, Dordrecht: D. Reidel, 1971.

Manketlow, K.I., and Evans, J.St.B.T(1979). Facilitation of reasoning by realism: effect or non-effect? *British Journal of Psychology*, 71: 227-231.

Manktelow, K.I. and Over, D.E. (1991). Social roles and utilities in reasoning with deontic conditional, Cognition, 39: 85-105.

Manktelow, K.T. and Over, D.E., eds. (1987). *Rationality: Psychological and Philosophical Perspectives.* London: Routledge.

Menger, Karl, (1939). "A logic of the doubtful: On optative and imperative logic," in *Reports of a Mathematical Colloquium*, 2nd series, 2nd issue, pages 53-64, Notre Dame, Indiana: Indiana University Press.

Oaksford, M. and Chater, N. (1994). A Rational Analysis of the Selection Task as Optimal Data Selection, Psychological Review 101(4): 608-631.

Oaksford, M. and Chater, N. (1995). Discussion: Information gain explains relevance with explains the selection task, *Cognition*, 57: 97-108.

Pollard, P. (1981). The effect of thematic content on the 'Wason selection task', 1: 21-29.

Popper, K. (1963). Conjectures and refutations: the growth of scientific knowledge. London: Routledge & K. Paul

Ross, K.L. (1996). The Foundations of Value, Part I: Logical Issues: Justification (quid facti), First Principles, and Socratic Method after Plato, Aristotle, Hume, Kant, Fries and Nelson, The *Proceedings of the Friesian School, Fourth Series*, published online at http://www.friesian.com.

Smith, T. C. and Witten, I. H. (1993). "Language inference from function words," Working Paper Series 1170-487X-1993/3, Department of Computer Science, University of Waikato, Hamilton, New Zealand, August 1993.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, 46(1-2):159–216.

Sperber, D., Cara, F., and Girotto, V. (1995). Relevance theory explains the selection task, *Cognition*, 57(3): 209-268.

Stenning, K. and van Lambalgen, M. (2001). Semantics as a foundation for psychology: A case study of the Wason selection task, In Journal of Logic, Language and Information, 10(3): 273-317.

Stenning, K. and van Lambalgen, M. (1999), Is psychology hard or impossible? Reflections on the conditional, In J. Gerbrandy, M. Marx, M. de Rijke, Y. Venema (Eds), Liber Amicorum for Johan van Bentham's 50th birthday, Amsterdam University, Amsterdam.

Stenning, K. and van Lambalgen, M. (2002), The natural history of hypotheses about the selection task: towards a philosophy of science for investigating human reasoning, In. Manktelow, K. and Chung, M. (Eds.) *Psychology of reasoning: historical and theoretical perspectives*. Psychology Press.

Stenning, K. and van Lambalgen, M. (2004) A little logic goes a long way: basing experiment on semantic theory in the cognitive science of conditional reasoning. Cognition, v. 24(4), pp. 481-530.

Stenning, K. (2002), Seeing Reason: Image and language in learning to think, Oxford; New York : Oxford University Press.

Stenning, K. and van Lambalgen, M. (submitted) A working memory model of relations between interpretation and reasoning. Submitted to Cognitive Science.

Landauer, T., and Dumais, S. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, v.104 pp. 211-240.

Vallduvi, E. and E. Engdahl (1996). The cross-linguistic realisation of information packaging. *Linguistics*, **34**, pp. 459-519.

van der Torre, L. (1999). Violation Contexts and Deontic Independence, *Proceedings of Context '99*, Trento, Italy, pages 361-374.

Wason, P, and Green, D. (1984). Reasoning and mental representation, *Quarterly Journal of Experimental Psychology*, 36(A): 598-611.

Wason, P. (1987). Problem solving. *Entry in the Oxford Companion to the Mind*, R. Gregory (eds.) pages 641-644

Wason, P.C. (1966). Reasoning, In B. M. Foss (Ed.) *New horizons in psychology*. Harmondsworth, UK; Penguin.

Wason, P.C. and Johnson-Laird, P.N. (1970). A conflict between selecting and evaluating information in an inferential task. *British Journal of Psychology*, 61(4): 509-515.

Wason, P.C. and Johnson-Laird, P.N. (1972). Psychology of Reasoning: Structure and content, Harvard University Press, Boston.

Wason, P.C. and Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, 23: 63-71.

Wason, P.C. (1968). Reasoning about a rule. Quarterly Journal of Experimental Psychology, 20: 273-281.

Wason, P.C. (1969). Regression in reasoning? *British Journal of Psychology*, 60: 471-480.

Wiemer-Hastings, P., Zipitria, I., (2001). Rules for syntax, vectors for semantics, Proceedings of the Twenty Third Annual Conference of the Cognitive Science Society, Mahwah, NJ: Lawrence Erlbaum Associates.

Wiemer-Hastings., P., (2000). Adding syntactic information to LSA, Proceedings of the Twenty Second Annual Conference of the Cognitive Science Society, Mahwah, NJ: Lawrence Erlbaum Associates, pages 989-993.