2003

EXPLORATION OF RELEVANCE EFFECTS IN REASONING

SIMON FRANCIS VENN

PHD

EXPLORATION OF RELEVANCE EFFECTS IN REASONING

VENN, SIMON FRANCIS

Award date:
2003

Awarding institution:
University of Plymouth

EXPLORATION OF RELEVANCE EFFECTS IN REASONING

by

SIMON FRANCIS VENN

A thesis submitted to the University of Plymouth in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

Faculty of Human Sciences

April 2003

Copyright Statement

Abstract

The study examines possible underlying mechanisms that may be responsible for generally observed biased response patterns in two conditional reasoning tasks: the Wason selection task and the conditional inference evaluation task. It is proposed that memory processes that may account for priming phenomenon, may also account for the phenomena of matching bias and double-negation effects in reasoning. A new *mental activation model* is proposed, based on distributed theories of memory, which models relevance effects of the problem materials by way of a simple algorithm. The model is seen to parsimoniously predict previous general response patterns found using the two reasoning tasks and makes unusual predictions concerning the size of the concepts used in the reasoning problems. The findings show that matching bias can occur between materials that do not lexically match but correlate on a semantic basis, which clarifies a previously uncertain area in the literature. It is also shown that previously deemed 'irrelevant' or mismatching cards on the selection task can interfere with the perceived relevance of matching cards if they are semantically related. The findings also show a weak but significant effect of concept size on matching bias in the inference task, supporting the proposed mental activation model. Issues concerning the notion of relevance perceptions being measured by particular response choices are raised with respect to both the selection and inference tasks.

Acknowledgement

I would like to thank Prof. J.St.B.T. Evans and Dr. I. Dennis for their supervision and helpful comments whilst carrying out the research and writing of the thesis.

I would also like to thank the Economic and Social Research Council (ESRC) for funding this research.

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

This study was financed by the Economic and Social Research Council (ESRC) award – R00429934451.

Signed: ....

Date: 26th August 2003

Contents

List of Tables and Figures

Tables

Figures

Chapter 1

Reasoning

Introduction

The thesis is an investigation into patterns of biased responses from deductive reasoning tasks. The deductive reasoning task of interest is one that involves conditional logic of the form, 'if something occurs then a consequent action will also'. People encountering these conditional statements, particularly in the form of abstract problems (i.e. letters, numbers, symbols), do not generally respond in a way that is logically correct according to the principles of formal logic. Instead, their erroneous responses have been found to fall into general patterns. These patterns of responses are closely related to the materials that make up the problem and are proposed by major contributors in the area, to stem from representational or interpretative anomalies. The present research investigates why the problem content may affect the way reasoners respond to conditional problems. More precisely, it will be hypothesised that the process of memory causes some materials in a problem to appear more relevant than others. Over the next three chapters, an argument will be developed which quite simply suggests that reasoning biases can occur because of short-term memory traces caused by reading the problem. That reasoners select or endorse certain answers to problems because of the concepts that were mentioned in the main premise.

A perception of relevance is proposed to emerge from the way memory is activated by the verbal or orthographic communication of the concepts mentioned within a problem and that these unconscious processes may lead to intuitions concerning the solution to the problem. Rather than see possible memory biasing effects as some sort of faulty processing, it will be proposed that they are merely artefacts of an efficient and powerful processing system, which under normal circumstances is well adapted to its environment.

What this study sets out to achieve is to answer as yet unresolved issues concerning the initial representational stage of reasoning.

It will be shown that when the content of a problem is represented in a distributed and associative memory network, rather than distinguish between current theories of reasoning it instead brings them together at a lower and more detailed level of description. Many theories of reasoning mention the effects of the initial representation and interpretation of the problem but none have yet to specify in any detail how these factors relate to current theoretical issues with memory. It would appear from current findings within the field of reasoning that the representational stage appears to have the largest effect on the patterns of responses from reasoning tasks. A more detailed account linking reasoning with distributed models of memory may show that reasoning biases are an emergent property of the way problem concepts are represented and activated in memory.

One thing that the mind does without debate, is to remember things – in fact all living things (animate at least) would seem to have a memory of sorts. Because it is such a widespread inter-species phenomena, we could assume that it is one of, if not *the*, most basic and primal cognitive functions. Without memory there would be no specific attention – we would merely attend to everything without filtering or attenuating the information. We couldn't *think*, as we would have nothing to mentally manipulate. It cannot be disputed that any 'higher order' cognitive functioning would have stemmed from the process of memory – a process that was already well in existence. The ability to reason must be reliant on the process of memory. Patterns of biased responses to deductive problem contents possibly transpire because of memory activation anomalies and difficulties arising from representing problem concepts such as negations. It may not necessarily be the mental manipulation of the *propositional calculus* that is being studied in reasoning tasks, but the behaviour elicited by the problem content and context materials.

This first chapter aims to give a broad outline of the current issues with respect to conditional reasoning. The general paradigms will be explained, followed by an account of the response biases that are seen to derive from the problems – primarily matching bias and effects of negation. The chapter will conclude with three of the main theoretical viewpoints concerning conditional reasoning – dual process, mental model and optimal data selection accounts. The details outlined in this chapter will be expanded on and considered in more depth in Chapter 3, wherein it will be discussed how memory issues might also be applied to the current findings in conditional reasoning problems.

Deduction

Research into the way people reason has primarily focussed on deductive competence. Essentially this involves creating problem formats where all the necessary information is given in order to select what has been preordained as a correct response. This is basically distinct from another form of reasoning, 'induction', where not all the information is present to form a correct response, but where reasoners have to fill in gaps with their background knowledge. The deduction paradigm is preferable because it is more transparent in that the information given is controlled (and not relied upon from another source) and the correct answer can be worked out according to the rules of a normative propositional logic (for a review see Evans, 2002).

The propositional logic of conditionals concerns four possible cases that can occur given a standard conditional premise such as 'if P then Q' (if *something* occurs then a *consequent* will also). These four conditions can be outlined in a *truth table* whereby the 'case' represents two occurring events in relation to the concepts mentioned in the conditional statement;

Case	Truth of rule
P,Q	True
P, not Q	False
not P, Q	True
not P, not Q	True

It can be seen that there is only one instance that can falsify the conditional statement – 'P, not Q'. All the other cases offer support for the conditional. It was suggested by Wason (1966) that the last two cases were psychologically implausible in that the absence of the antecedent clause (not P) would be seen as irrelevant to the conditional. That is, given the premise, 'if it is a dog then it is an animal' it would seem irrelevant to the rule to be given the inference, 'there is a giraffe therefore it is an animal' (not P, Q) or that 'there is a chair therefore it is not an animal' (not P, not Q). Seeing the latter two cases as irrelevant is seen to be possessing a *defective* truth table and a departure is seen away from standard pure logical constructs towards a more psychological understanding of conditionals.

Some conditional statements are more refutable than others and can go from general experience – 'if I put my hand in the fire then it will hurt' to more scientific hypothesis testing – 'if the two chemicals are mixed together, then they will turn green'. Being able to manipulate the inferences correctly – that is according to the principles of normative logic – will enable a better understanding of their truth. For example, with the chemical inference above, it would be fallacious to presume the two chemicals had been mixed together had the substance turned green. It may have done so on its own with time or with atmospheric conditions for instance. Similarly, if my hand hurts, it does not necessarily follow that I had put it in the fire – I could have hit it accidentally. These would be fallacious inferences and highlight the directionality of the logic required to correctly manipulate the information contained within the conditional. To accept an inference as the one above, would imply a bi-conditional of the form 'if P then Q' *and* 'if Q then P'.

Reasoners generally see both 'P' and 'Q' as important to ascertaining the truth or falsity of a conditional statement. But, the presence of 'Q' has no logical bearing on 'P' unless the conditional is interpreted as a bi-conditional The existence of 'Q' does not necessarily imply the presence of 'P'.

Conditional Problem Formats

Truth Table Task

There are several forms of conditional problem formats; truth table task, selection task and conditional inference task. A truth table *evaluation* task was used by Johnson-Laird and Tagart (1969) in which participants were told that cards had a number and a letter printed on them. They were then given a conditional premise of the form, 'if the letter is a B then the number is a 7' and were asked to evaluate the cards as either true, false or irrelevant. Their findings supported the contention that people possess a defective truth table in that cards which did not have the antecedent clause printed on them were seen as irrelevant. There has also been a *construction* version of the truth table task in which participants had to construct a letter/number pairing to either verify or falsify the given conditional. Results from the construction tasks coincide with those from the evaluation tasks (Evans, 1975) and also reflect the notion of a defective truth table in that letter/number pairings in which the antecedent was not represented, were not constructed.

Wason Selection Task

There have generally been two forms of the Wason selection task (Wason, 1966; for a review see Evans, Newstead & Byrne, 1993); one using abstract materials (letters, numbers, symbols) and one using thematic or real world content (usually with a scenario). In the abstract version, participants are typically shown four cards e.g. depicting A, K, 4, 7 (represented in the propositional calculus by P, not P, Q, not Q respectively). They are informed that cards have a letter on one side and a number on the other, and are given a conditional premise in the guise of a rule, such as, 'If there is an A on one side, there is a 4 on the other side'. Their task is then to select cards to turn over which could decide if the rule is true or false. As the rule can only be falsified by an A card without a number 4 on the back, the logically correct selection would be the A and 7 cards. Most participants get this wrong, choosing predominantly the A card or the A and 4 cards. The general pattern of choices made in the selection task are typified by Oaksford and Chater's (1994, Table 2) meta-analysis of 13 studies which show a mean proportion of cards selected as; .89(P), .16(not P), .62(Q), .25(not Q). The pattern of cards generally selected seems to indicate that reasoners see both 'P' and 'Q' as important to the verification of the conditional statement, and might also indicate that reasoners are using a bi-conditional definition of the conditional.

Conditional Inference Task

The conditional inference task again involves giving participants a conditional statement, but this time it is followed by a minor premise concerning either the antecedent or consequent clause. Reasoners are then either required to produce their own conclusion in the *production* version of the task, or to evaluate a given conclusion in the *evaluation* version. A typical example would involve the use of a *major premise* such as 'if the animal has four legs then it is a dog'. To form a typical problem format, this statement would then be followed by a *minor premise* such as 'the animal does not have four legs'. Reasoners are then either required to produce their own *conclusion* – is it a dog or not, or they are asked to evaluate a conclusion that is given to them e.g. 'it is not a dog' – *does this follow or not?*

Typically only four combinations of *major premise*, *minor premise* and *conclusion* have been of interest. These combinations come under the headings of *modus ponens* (MP), *modus tollens* (MT), *denial of the antecedent* (DA), and *affirmation of the consequent*

(AC). The latter two are classed as logically fallacious unless a bi-conditional interpretation of the conditional is taking place. Examples of these combinations are:

MP – If there is a P then there is a Q. There is a P. Therefore there is a Q.

MT - If there is a P then there is a Q. There is NOT a Q. Therefore there is NOT a P.

DA - If there is a P then there is a Q. There is NOT a P. Therefore there is NOT a Q.

AC - If there is a P then there is a Q. There is a Q. Therefore there is a P.

Negations Paradigm

Negations can be systematically introduced into the conditional premise to provide four major premise types i.e. *if P then Q, if P then NOT Q, if NOT P then Q* and *if NOT P then NOT Q* and subsequently the polarity of the minor premises and conclusions changes accordingly. This methodology was originally used by Evans (1984) and has been labelled the 'negations paradigm' by Oaksford and Stenning (1992). So, for example if we take the MP inference and systematically introduce the negations paradigm, the minor premise and conclusion have to change polarity accordingly to remain logically equivalent, such that:

If there is a P then there is a Q. There is a P. Therefore there is a Q.

If there is a P then there is NOT a Q. There is a P. Therefore there is NOT a Q.

If there is NOT a P then there is a Q. There is NOT a P. Therefore there is a Q.

If there is NOT a P then there is NOT a Q. There is NOT a P. Therefore there is NOT a Q.

(See Appendix A for a full list of the 16 possible inferences)

Matching Bias

One of the context biasing effects that is prevalent in the reasoning literature is the phenomenon of *matching bias* (see Evans, 1998 for a detailed review). Matching bias was first recognised by Evans (1972) and Evans and Lynch (1972) following up Wason's

(1965;1968) investigations into verification bias. Evans found that by introducing negations into the conditional premise, reasoners could be manipulated into giving logically correct responses. Using the standard conditional 'if P then Q' in the selection task, Wason found that reasoners would predominately choose the P and Q cards (the logically correct choices here being the P and not Q cards). By inserting negations into the premise i.e. if P then not Q, Evans found that reasoners still predominately gave the same response of P and Q, albeit the correct response pattern now. It seemed that the negation had been ignored and what Wason had believed to be verification or confirmation bias was in fact just a tendency to choose a response item that *matched* the items mentioned in the premise.

The matching bias predictions for the selection task can also be applied to the conditional inference task. Instead of being more inclined to choose cards that match in the selection task, the conditional inference task can highlight matching bias by a greater tendency to *endorse* more conditional inference conclusions in which the minor premise item matches the item mentioned in the major premise. There are specific circumstances involving negations in the problem, which are necessary precursors to causing the matching bias effect. Specifically, the matching bias effect is only seen to occur with what have been labelled 'implicit negations'. Details on these factors will now be given in the following section.

Conditions for Matching Bias

Evans (1998) outlines three factors affecting matching bias: Linguistic generality, realistic/thematic material, and implicit negation. These three aspects provide a useful structure in discussing the matching bias phenomenon and show how problem materials can contribute to, or diminish the matching bias effect.

Linguistic Generality

It had been thought that the matching effect was peculiar to conditional 'if' statements. This was substantiated in several forms of conditional inferences, not only 'if…then' conditionals but also 'only…if' (Evans 1975) and reverse 'if' (i.e. Q if P) conditionals (Evans, Clibbens & Rood, 1996) on both truth table tasks and the Wason selection task. Evans and Newstead (1980), Evans (1989), and Oaksford and Stenning (1992) therefore believed the evidence indicated that matching bias was only observable using conditional inferences. More complete evidence has since been obtained by Evans, Legrenzi and Girotto (1997) who studied the responses to not only conditional statements but also 'universal' (e.g. every P has a Q), 'disjunctive' (e.g. either a P or a Q) and 'negative conjunctive' (e.g. it is not possible to have a P and a Q) logical statements. Evans (1998) summarises, "The suggestion is that matching is a phenomenon locally connected to negations within rules and not requiring the linguistic context set by the word 'if' or even a semantically directional expression of implication". Evans does not imply by this that matching only occurs in the presence of negations, but on affirmative concepts also. Matching bias would seem therefore to be elicited during most logical problem solving operations that rely on linguistic presentations or formats.

Being such a universal phenomenon suggests that it occurs due to the basic representation of the problem (instigated by the linguistic materials) before an attempt is made at mentally trying to solve the problem in accordance with the written instructions (and subsequent logical interpretation). As Evans (1998) points out, only two present theories can explain such universality – the heuristic-analytic theory and the processing negations account of Oaksford and Stenning (1992). Only the two theories concentrate on the actual representation of the materials and not any attempt at logical problem solving or mental manipulation of the materials.

Realistic/Thematic Material

It will be remembered that abstract reasoning tasks involve the use of letters, numbers, symbols etc. to construct the problem e.g. 'if there is an A on one side then there is a 2 on the other'. Thematic problems on the other hand utilise real-world concepts set in a realistic context. There are many studies in the literature that outline patterns of responses from deductive tasks which indicate there is more affecting responses than just the logical structure of the problem itself or even problems of interpretation or representation of the negations. For instance, there have been studies investigating realistic rather than abstract materials (examples of which will be given below). These studies have claimed a facilitatory effect on logical responses using real world scenarios. This may be because abstract tasks are too removed from the way people reason normally and are naturally therefore prone to errors and biases.

The effects of thematic materials on matching bias are of particular interest to the current study. It will be seen that there are circumstances using thematic materials when matching bias remains at similar levels to that seen in abstract tasks. There are other circumstances using thematic materials when matching bias diminishes or even disappears altogether. Some of the current theoretical explanations for this may have direct relevance to theories of a semantically organised and distributed system of memory.

Three sets of studies using the selection task will be outlined that have used thematic materials. They generally show a chronological order of investigation into apparent effects of thematic materials in conditional inference tasks and highlight how the area has developed. The first set of studies showed that matching bias can diminish and logical facilitation increase because of a *familiar* context. The second set of studies also show a decrease in matching bias but also a decrease in logical responding. The final study shows that matching bias can still be observed using thematic materials and that thematic effects on matching bias may be due to the presence or absence of a comprehensible context in which the materials are placed.

Familiar Context

Griggs and Cox (1982) used a drinking-age law scenario with selection task cards depicting ages and beverages. It was felt that such drinking age laws were familiar to most participants as they tend to exist in most countries. They used a rule, 'If a person is drinking beer then that person must be over 19 years of age'. The selection task cards depicted, 'beer', 'coke', '16 years of age' and '22 years of age'. A short scenario was given to participants at the beginning of the task in which they were to assume that they were police officers who were checking in a bar to see if there were any under-aged drinkers. The results from the study showed that with a short scenario, participants were more likely to select the correct 'P' (beer) and 'not Q' (16 years of age) responses. Hence the matching bias effect diminished. Without the short scenario, logical facilitation decreased.

Similar facilitatory findings were observed for Johnson-Laird, Legrenzi and Legrenzi (1972) who used a postal rule concerning envelopes and costs of postage (a rule which was familiar to the participants). The rule here was, 'If the letter is sealed then it has a 50 lire stamp on it'. The cards depicted pictures of, 'a sealed envelope', 'an unsealed envelope', '50 lire stamp' and a '40 lire stamp'. The scenario informed participants that they were postal workers checking the validity of the postage. Again, similar logical facilitation was observed as was the case in the 'drinking-age rule'. It was felt that these studies highlighted the fact that familiar and easy-to-understand contexts enabled participants to reason logically and that biases from abstract tasks were caused by their lack of real-world significance.

Further support for the *familiarity* hypothesis comes from a follow-up study of Cheng and Holyoak (1985) who repeated Johnson-Laird, Legrenzi and Legrenzi's (1972) postal study but on populations who had not encountered this rather unique postal rule. The pattern of responses obtained from these studies was more akin to those found using abstract tasks. It was felt that the delineating factor between the studies could only have been the previous experience and memory of the postal rule and not just the use of thematic materials themselves.

Novel Context

Evans (1995) used realistic materials with and without a clarifying context in a selection task. In Evans' Experiment 2, arbitrarily thematic contexts were used in a selection task – that is, materials in which the scenario was one that would not have been encountered before (unlike Griggs & Cox's, 1982, 'drinking age rule') but which made sense having been given a real world context. Such a scenario was deemed to lack the facilitatory pragmatic cues that could have increased logical responding on tasks such as the drinking age rule. An example of the scenario used by Evans is:

> *You visit a friend of yours, Sarah, who is an art student. "I have just been working on my Christmas cards", she says. "I thought I would economise this year by making my own cards with my friend's names written on one side and an illustration of my own on the back". You look at the four cards she has finished.*
> *"I have decided that it would be a good idea to stick to the following rule: If a card has [does NOT have] a male name on one side then it has [does NOT have] a red flower on the other side".*

With four cards depicting; 'Paul', 'Mary', 'picture of a red flower' and a 'picture of a yellow flower'.

Two variables were used in the experiment; the presence or absence of the scenario; and reasoning or judgement instructions. Reasoning instructions were the standard selection task ones where reasoners were to ascertain which cards needed to be turned over to see if the given rule is true or false. Judgement instructions asked reasoners to decide which of the four cards 'appears to be relevant' to the rule. The findings showed that matching bias on the reasoning task was eliminated when a scenario was present but was elicited when a scenario was not present. Overall logical facilitation was not increased by the presence of the scenario. It would seem that novel thematic materials reduced matching bias in the standard selection task only when a scenario was present but did not facilitate logically correct responses.

Abstract Thematic Context

It would seem that a scenario, be it a reflection of a real world situation or a novel scenario reduces or even eliminates matching bias. The exception to this general finding is Manktelow and Evans' (1979) finding that matching bias is still present when a scenario is used. They used rules such as, 'if I eat haddock then I drink gin' with four cards depicting food and drink. The scenario was brief in that it just told participants what food and drink combinations were being consumed at particular meals. The findings using these thematic materials did not significantly differ from those using abstract materials. It would seem that thematic materials alone are insufficient to reduce matching bias or facilitate logical responding. It may be, as Evans (1995) points out, that the content of the Manktelow and Evans' study was so 'semantically impoverished' that it could really be classed as a standard abstract task. There is much debate about the causes of these selection differences using thematic materials (see Evans & Over, 1996, Chapter 4 for a review). It is more pertinent to the present discussion to now pick up on one main theme that has arisen – that of the pragmatic nature of the tasks.

Pragmatics

Explanations for the thematic/abstract phenomena have concentrated on the pragmatic information offered by the context of the problem. It has been argued that realistic materials invoke a set of pre-learned rules for obligations and permissions (Holyoak & Cheng, 1995). They claim that a thematic reasoning problem elicits pragmatically cued schemas – previous knowledge of rules that are seen as being relevant to a new problem. A problem is then 'mapped' on to one of these rules by the pragmatic context given by the problem. The idea of using innate or pre-learned rules depending on the context though seems a little vague in that it is not clear what aspects of the problem are needed in order to invoke a schema and how the correct schema is selected.

Evans and Clibbens (1995) claim that it is not pragmatic schemas that are invoked by such materials, but the notion *of pragmatic relevance*. The pragmatics within the scenario or context cues certain materials in the problem as appearing more relevant than other materials. By cueing other non-matching materials, matching bias may be reduced but logical responses may not necessarily increase. Evans (1998) believes that matching responses are more likely to occur when other pragmatic cues to relevance are weak. The idea of cues or pragmatic relevance in these respects could be linked to the way these concepts and scenarios are organised in memory.

Pragmatics are seen as a form of implied information that evokes previous knowledge of contexts and situations, in addition to the literal truth offered by an utterance. Pragmatics are dependent on the context in which the utterance was mentioned. They are in effect another form of meaning, just as the semantics associated with words. Evans (1989) makes the distinction between linguistically cued relevance and pragmatically cued relevance to describe heuristic effects found on abstract and thematic conditional tasks respectively. Essentially in both instances the focus is on the meaning and the semantics associated with verbal communication, but presented in different ways. They both involve the mental associations that are activated with a verbal communication. The only difference being the quantity of associations that are related to the communication. This is recognised by Evans (1995) who described the thematic materials used by Manktelow and Evans (1979) as 'semantically impoverished'. Enriched, thematic materials which reduce matching bias have more semantic associations (both of a pragmatic and lexical nature). Abstract content has far fewer associations and evokes matching bias. It would seem then, there is a direct relationship between the amount of memory activation evoked by a problem and the amount of matching bias elicited.

Memory Representation

What these studies point to is the issue of how the different materials used can affect responses to what are logically the same tasks. The materials do not change the logical requirements of the task so they must change the way reasoners mentally represent them. The only difference between an abstract task and a thematic one, are the associations people have with the materials. In effect, the manipulation of thematic/abstract content could be thought of as a memory manipulation, not a reasoning one. Logical facilitation occurs only when some form of transferable experience can be related to the rule. Matching bias on the other hand only seems to occur with semantically impoverished materials.

What is predominately interesting about these studies is the interaction of memory with the reasoning tasks. Manktelow and Evans (1979) commented that facilitatory effects of using realistic materials might be due to 'memory cueing effects' rather than just using concrete materials themselves. The general conclusion of their research was that realistic materials need a comprehensible and realistic context (usually in the form of a scenario) if they are to change the pattern of responses from the normal pattern found using abstract tasks. Such a context may also aid memory by enriching the problem at hand with more information. Enrichment here could take the form of extended semantic activation in memory.

Stanovich (1999) uses the umbrella term 'fundamental computational bias' for the biases caused by previous memories, wherein the problem is contextualised with regard to previous experience and knowledge. A problem may be contextualised from having greater associations with the materials. The mental activation of all these related concepts seems to reduce the relevance attributed to certain materials more than others (which may have been the initial cause of matching bias). Either it may be that relevance of all the items mentioned in the problem increased, so that no one in particular stands out, or that the relevance to all items is decreased due to the amount of mental activation that is taking place. Whichever way it is looked at, it would seem that matching bias would be caused by *differences* in perceived relevance of the materials and if these differences are decreased then it might be expected that matching bias would also.

The results from tasks using thematic materials are offered as support for the contention that differences in the levels of mental activation in representing the concepts in a problem are directly responsible for matching bias. The larger the spread of activation (caused by semantically enriched concepts) the less likely the problem is to elicit matching bias. If the problem does not evoke such a distributed amount of mental activation (abstract tasks), but instead activation is solely concentrated on the four concepts mentioned in the selection task for instance, then matching bias is more likely to occur.

Further, more recent established links between the amount of memory activation and performance on reasoning tasks comes from Markovits, Doyon and Simoneau (2002). They found a direct relationship between individual differences in the type of working memory capacity and logical performance on conditional reasoning tasks. They divided working memory measures into verbal and visual working memory measures using both abstract and real world materials. Whilst verbal memory was correlated with logical performance on both abstract and thematic tasks, the visual memory measures only correlated with logical performance on the thematic tasks.

The findings indicate a direct link between the type of mental representation required by the problem materials and logical responses to conditional reasoning tasks. Thematic materials would seem to require, or evoke both verbal and visual memory. Whereas, performance on reasoning tasks with abstract content relies on verbal working memory. It would seem that thematic materials evoke an extra form of mental representation – visual working memory. Participants with good visual working memory are more likely to solve thematic problems than abstract ones. It could be this evocation of visual working memory that accounts for the embellishment or enrichment associated with thematic material. Embellishment stems from activating more mental associations (in this case visual ones) and this additional activation reduces bias and can help solve the problems.

Implicit Negation

Finally, the last factor of the three that is seen to affect matching bias is implicit negation. Using the four premises of the negations paradigm (alternating polarities), the general finding is a matching bias for the named premise item irrespective of negation. This not only happens in the selection task but also in the conditional inference task where reasoners endorse or construct a conclusion rather than select cards. The negations used in the conditional problem have to be *implicit* to induce the matching bias effect (Evans, Clibbens and Rood, 1996). A problem would be deemed as providing implicit negations, when for instance the possible response choice (selection task) or minor premise to be endorsed (inference task) contains an item which is 'not P' or 'not Q' without explicitly stating the negation. So, using the example, 'if there is a vowel on one side, then there is not an even number on the other side', the card or minor premise could have the implicitly negated consequent of say '3' – which is implicitly 'not an even number'. In tasks using *explicit* negations, the card or minor premise would be literally 'not an even number' – a direct copy of the clause in the major premise. When explicit negations are used in both the selection task and the conditional inference task, the matching bias effect disappears.

The following are examples of implicit and explicit negations in both the selection task and inference task:

Selection Task

'If there is an A on one side then there is not a 2 on the other'

Implicitly negated cards: 'A', 'B', '2', '3'

Explicitly negated cards: 'A', 'not A', '2', 'not 2'

Inference Task (Affirmation of the Consequent)

'If there is an A on the left side then there is not a 2 on the right side'

Implicitly negated minor premise: 'There is a 3 on the right side'

Explicitly negated minor premise: 'There is not a 2 on the right side'

Conclusion: 'Therefore there is an A on the left side'

It can be seen that the matching bias effect occurs in the selection task between the *card/premise clause* pair and in the inference task between the *minor premise/major premise* pair. Controlling for the logical requirement of the task, matching bias is normally observed using implicit negations in these pairs when there are more 'matching' '2' / '2' and '2' / 'not 2' selections (or inference endorsements) than 'mismatching' '3' / 'not 2' and '3' / '2' ones.

Using explicit negations involve substituting the clause 'not 2' instead of '3'. By doing this it can be seen that the topic of the clause (irrespective of negation), now matches between the major premise and card/minor premise i.e. 'not 2' / 'not 2' or 'not 2' / '2'. By keeping the topic the same on both sides, it can be seen that matching of the terms could now be taking place.

Oaksford (2002) has suggested the idea of two levels of bias – $matching_1$ and $matching_2$ which describes how ignoring the negation in a conditional inference task and matching the named item (the usual form of matching bias) can be classed as *matching*$_1$. But in the case where explicit negation is used (either as a card choice or as a minor premise) a match of the *whole clause* could be taking place (*matching*$_2$) e.g. a match could happen between 'not Q' in the referred clause to 'not Q' in the categorical premise.. This would create the 'illusion' of a reduction of the standard matching responses (*matching*$_1$) because more than the usual amount of negated clauses would be chosen. Oaksford (2002) chooses to see this second type of matching as a strategy rather than a 'psychologically interesting reason'.

It seems apparent that this simple explanation for such behaviour during a reasoning task reflects the fact that participants are swayed by the linguistic representation of the materials and concepts, than they are with the logical content or negations held within the problem. So, although both the implicit and explicit negations can be seen to be the same logically, they differ in the way they visibly or even semantically match (or mismatch) the items mentioned in the major premise.

There are two aspects with regard to memory and the way negations are displayed in a conditional problem. First, with regard to implicit negations it can be seen that an implicit negation would not activate the same (or part of) semantic network that is activated by the concept mentioned in the initial premise. Using the inference task as an example it can be seen that they are no longer the same concepts i.e. given the premise, 'There is an *A* on the left side therefore...', an implicit minor premise might state, 'There is a *B* on the left side.' In terms of mental representation, the activation of the concepts no longer 'overlap'. That is, they do not activate exactly the same area as they would have, had the minor premise been, 'There is a *A* on the left side'. It will be argued in this study that it is this overlap in mental activation that causes a heightened perception of relevance. Implicitly negated concepts (mismatching) are not seen as relevant because they do not overlap with the item mentioned in the initial premise to the same extent as the matching item does.

Secondly, it is argued that an explicitly negated concept (e.g. 'not a cat') can only be represented in memory by activation of the mentioned item (e.g. 'cat') - but perhaps at an inhibited level of activation. Logically the clause 'not a cat' represents *everything* but a cat, but it will be argued in Chapter 3 that this is psychologically implausible in terms of memory activation. The notion that an explicitly negated concept is itself activated rather than all the concepts it represents could explain why explicitly negated items are still seen as being relevant to the problem. The activations of both the major premise item (e.g. 'cat') and explicitly negated minor premise item (e.g. 'not a cat') overlap in their activation of the same semantic network.

Theories

Heuristic-Analytic & Dual Process

Evans (1984;1989) gave an 'heuristic-analytic' account of reasoning which incorporated the idea that there were two psychological factors at work when encountering a reasoning problem. This 'dual process' theory of reasoning was later more clearly formulated (Evans & Over, 1996) and explained how apparently illogical biased responses (matching bias) could be given on a problem such as the selection task and how such biases were more a reflection of an adaptive system than they were 'illogical'. Evans (2002) has recently linked the earlier dual process account (Evans & Over, 1996) with Stanovich's (1999) *System1* and *System2* processes. This new terminology neatly incorporates other general cognitive functioning dichotomies including Reber's (1993) *implicit* and *explicit* learning distinction and Sloman's (1996) *associative* and *rule* based systems. System1 processes describe how biases can occur due to the problem content itself and can explain how matching bias can occur because of initial unconscious selections of what are perceived to be relevant aspects of the task.

System1 processes are described as fast, unconscious, parallel processes (as opposed to slower sequential conscious System2 processes). Such processes are paramount to creating the initial representation of a problem. System1 is described as being represented as a collection of powerful localised neural networks that operate almost independently of each other. Evans sees perceptions of relevance as the goal of the System1 as it constructs a mental representation of the problem. System2 processes are more closely related to the kind of thought required for formal propositional logic and are classed as a general purpose cognitive ability to manipulate the relevant information given forth by System1. Links are made with this general System2 ability and IQ. This hypothetical-thinking characteristic is also said to represent information as well as manipulate it but it does not seem parsimonious to have the representation of concepts in both System1 and System2. It would be more efficient to have a separate representation and manipulation system, which would place System1, as very much a perception and memory process operated on by System2 processing. System2 processes are deemed to be more closely related to individual differences in general intelligence (Stanovich & West, 2000), relying on cognitive facets such as working memory (Markovits, Doyon & Simoneau, 2002) whereas System1 processes are universal, with an implication that they are in some way more primordial. Thus relevance effects caused by a System1 problem representations would always restrict further analytic processing, as they create the initial stage for any further conscious thought.

Evans (1998) uses the *heuristic-analytic* (HA) theory to explain the matching bias phenomenon because the heuristic elements (the 'if' and 'matching' heuristics) are more specific to matching bias, although the HA theory still reflects the general duality of cognitive performance with System1 and System2 processes. The 'if' heuristic describes how true antecedent (TA) responses are generally preferred to false antecedent (FA) responses because of the linguistic nature of the word 'if' directing attention to detecting the presence of the 'if' topic. The 'matching' heuristic describes how pre-conscious or pre-attentive processes direct attention to an item irrespective of negation. This focused attention heightens the perception of relevance of that item which restricts any conscious logical analysis because only the 'relevant' information is chosen to represent the aspects of the problem. Originally, the matching heuristic was described with regard to linguistic processes using Wason's (1965) proposal that negations are used to deny presuppositions and not to assert new information. Evans does not offer any further specificity for the mechanisms that create pre-attentive focussing.

This two-system approach applied to reasoning basically describes an initial, unconscious representation of the problem materials. Within this unconscious processing, some items of the problem are somehow perceived to be more relevant than other items. These relevant items then form the mental representation of the problem. Any subsequent analytic processes applied to the problem are applied to this unconsciously formed (System1) representation. Therefore, biased responses to the problem content are deemed to stem from biases that have been formed in the representation stage rather than the analytic or formal logic stage of System2 processing. System1 processing is seen as unconscious and therefore it must be presumed that reasoners have no control over what are perceived to be relevant aspects of the problem.

It could be argued that relevance judgements at this level are more likely to stem from mental architecture and the way memory is organised with respect to the items mentioned. At this level it may be more useful to think of System1 as a *mental activation* stage rather than suggest any form of processing. The notion of mental activation encapsulates the immediacy of what is taking place, placing it at a sensory or perceptual level. The notion of *processing* implies that some filtering, selecting or organising of the problem concepts is happening whereas *mental activation* implies an almost autonomic response to the stimuli.

Any type of processing at this System1 level can not be seen to be taking place. Otherwise, the implication would be that there is a form of *unconscious* analytic System1 that is more powerful than the *conscious* formal analytic System2. But any unconscious analytic processing system would be seen to be flawed in that it is *selecting* the wrong materials for the conscious system. The perception of heightened relevance of some items over others is automatic and must therefore reflect implicit memory activation anomalies, which are seen to be immediate and powerful parallel processes, rather than any form of unconscious processing or selection. In Chapter 3, it will be argued that System1 processes are essentially caused by the activation of a distributed memory network and that operating or processing biases stem from anomalies of trace activations – strongly linking the biases to the phenomenon of *priming*.

Mental Model Theory

The mental model approach of Johnson-Laird (1983) describes how information from a problem is represented, and how attempts are made at manipulating this representation and constructing falsifying cases. Biased responses are said to emanate from representational inaccuracies or working memory limitations in processing alternative solutions to the problem. Johnson-Laird (1983) and Johnson-Laird and Byrne (1991) proposed a theory of deduction based on the idea of mental models in which the concepts of a problem are represented in an abstract way with markers representing certain propositions. Thus a conditional problem would be represented in a fashion akin to an internal algebraic representation where aspects of the problem such as negation, exhaustivity and further possibilities are somehow 'flagged'. According to Johnson-Laird's theory, the standard modus ponens conditional (if P then Q) is represented as:

[P] Q

...

Where the parentheses '[ ]' represent the exhaustivity of that concept – here, *every* 'P' will have a 'Q'. The three dots (...) represent further possibilities yet to be 'fleshed out' or elaborated on. The model theory proposes that possible states of the world are represented mentally by the use of models that incorporate markers to represent propositions and their relation to each other. The mental model approach relies on a *simple semantic principle*. That is, a conclusion is valid if there are no models of the premises that exclude it.

The mental model construction is believed to go through the following stages: An initial explicit model is represented of the premise. The items in the model are ones that are seen to be relevant to the premise. The model is then used to generate provisional conclusions. Attempts are then made at validating these conclusions by generating alternative models of the premise and seeing if the conclusions still hold. If an alternative model can not be generated which excludes the provisional conclusion, then the conclusion is declared valid.

Deduction then, according to the mental models approach involves three stages (Johnson-Laird & Byrne, 1991)

1. Comprehension – Reasoners use previous background knowledge to understand the premises. They then create an internal model that represents the premises.
2. Description – "Reasoners try to formulate a parsimonious description of the models they have constructed". This description should contain information that isn't explicitly stated in the premise – like an algorithmic overview – basically the essence of the premise. If such a 'conclusion' can not be formed, then it is deemed that nothing follows from the premises.

3. Validation – If a tentative conclusion is reached from the previous stage, then alternative models of the premises are searched for that could make this assumed conclusion false. If such a model is not found, the conclusion is seen as valid. If a falsifying model is found, then there may be an attempt to return to the second stage to see if there is an alternative tentative conclusion that may be true for all the models constructed so far.

Johnson-Laird and Byrne state that this process could continue until all possible models are exhausted. Only in Stage 3 is there any real deductive work being carried out. The first two stages are merely comprehension and representation of the problem. It is not entirely clear how these models are represented in memory. Johnson-Laird and Byrne (1991) point out that, "The tokens of mental models may occur in a visual image, or they may not be directly accessible to consciousness. What matters is not the phenomenal experience, but the structure of the models" (Pg. 39). A couple of representational points that they make are of interest here. One concerns the representation of the negated constituent. They claim such negated clauses "elicit representations of the corresponding positive items" – a point they proffer for causing matching bias. In fact, they postulate that it is the initial representational stage, especially with 'neutral' conditionals, that gives forth biased responses, in conjunction with working memory limitations.

Evans and Handley (1999) have since refuted that explicit representations of the negated item (e.g. both 'not 2' and '2') are also represented in the initial model. As will be seen in Chapter 3, such an account has difficulties explaining the so-called 'double-negation' effect. Evans and Handley suggested that the mental model account be amended by not including a representation of the explicitly negated item. Johnson-Laird and Byrne (2002) have subsequently accepted this amendment, suggesting now that the matching bias phenomenon is caused by implicitly negated 'mismatching' terms in the representation of the problem.

It is this initial explicit representational stage that Evans (1991; 1993) has suggested coincides with the heuristic aspect of the HA theory of matching bias. Here, unconscious processes select aspects of the problem for representation (and possible further analysis) by their perceived relevance. In fact, Evans (2002) claims that this initial representation stage in the mental model theory is in itself part of System1 processes. Legrenzi, Girotto and Johnson-Laird (1993) have discussed the notion of *focussing effects* within the representational stage of forming mental models. This idea is parallel with Evans' notion of relevance in that model generation (true analytic reasoning) is restricted to what is explicitly represented in the problem. Subsequent alternative model generation is more akin to an explicit and conscious System2 process.

Another representational point of interest concerns the representational descriptions for syllogistic problems. Here they point out that representations of the problem can be modelled using Euler circles, or better still, Venn diagrams. Citing Johnson-Laird (1983) they claim, "A uniform and more powerful principle, however, is that mental models have the same structure as human conceptions of the situations they represent". This notion of representation here is important for the present study for the fact that they are claiming a direct link between problem representation and memory, and therefore inferring an indirect link between memory and biased responses. It is also pointed out that direct mental representations such as these would be hard pressed to represent negations. These points will be taken up in Chapter 3, wherein an attempt will be made at creating representations of these mental models of conditional problems using similar schematic diagrams to the Venn diagrams used by Johnson-Laird and Byrne (1991) to describe basic syllogisms.

Again, the mental model approach suggests representational anomalies of the problem content being a major contributor to causing biased responses. Although the representational process isn't fully described, it would appear as in the dual process account, that unconscious selection processes are taking place, which form the mental model. It has not been stated as to how negated concepts are represented in memory. A negated concept is unlikely to be represented solely by the activation of all the concepts in the negated set i.e. 'not a cat' is unlikely to be represented by the activation of all other concepts in memory. In fact, the representation of the negated concept has little relevance unless the item (without negation) is somehow represented. If the original concept was not represented in memory it would be difficult to remember or ascertain just what the activation of the other concepts represented. Oaksford and Stenning (1992) have offered a more detailed account of how negations are represented or interpreted.

Processing Negations & Optimal Data Selection

Oaksford and Stenning (1992) proposed that biased responses (particularly matching bias) occur in both the conditional inference construction task (their Exp. 1) and the selection task only when, "insufficient or ambiguous information prevents the intended interpretation of negations". The bias occurs because of a difficulty in processing negations. Difficulty in processing negations is proposed to stem from a difficulty in constructing the *contrast class*, which represents the negated clause. Here, because a negation is deemed as a psychological concept (Wason, 1965) used to deny a presupposition (and not necessarily a logical construct which implies 'everything but' the negated constituent) the contrast class is usually semantically linked in someway to the concept mentioned. That is, 'not a cat' might infer 'dog', 'budgie', 'snake', 'mouse' etc. hence the contrast set for 'not a cat' might be the category 'all other domestic animals'.

Although Oaksford and Stenning do not directly prescribe the contrast class as being semantically related to the negated concept, they do state that its construction is dependent on *world knowledge* and that the contrast class member should be as similar as possible. They state that a negated constituent could be *interpreted* from information from several implicit and unconscious domains such as phonetics, syntactics, semantics or pragmatics. Abstract tasks are seen as causing difficulty/biases with constructing contrast classes because they do not rely on or inspire the input of world knowledge. Also, in the case of the selection task, Oaksford and Stenning claim the instructions may lead to the ambiguous state whereby all three other card choices could represent 'not P'. That is, reasoners may confuse the antecedent and consequent selections.

Oaksford and Chater (1994) went on to give a 'rational analysis' of the selection task which is based on Bayesian probability theory. Here, they claim cards are chosen in order to maximise *optimal data selection*. Looking at rationality from an optimality point of view, which sees rationality as adaptive to the environment an organism is in, they do not follow the view that reasoning responses must resemble laws of normative logic in order to be classed as rational. Instead, a rational analysis view adopts the perspective that choices in reasoning tasks stem from attempts at optimising the expected amount of information given by that choice. They combine this approach with what they call the *rarity assumption* in order to explain the choices made in the selection task. The rarity assumption according to Oaksford and Chater assumes that, "P and Q are rare by default and that experimental manipulations influence these parameters by moving them away from their default values" (Pg. 625).

Oaksford (2002) explains how negated concepts represent high probability contrast sets. Using the example that the set 'drinks that are not coffee' is a lot larger than 'drinks that are coffee'. If something is a drink, the probability of it not being coffee is considerably higher than the probability of it being coffee. The lower probability set is deemed as more informative and is more likely to be a preferred card choice in the selection task. This can explain why given the premise, 'if A then not 2' participants are likely to appear to be matching by selecting the '2' card. In the set of 'single numbers' the '2' card represents the lower probability of '1 in 10' whereas 'not a 2' represents the probability of '9 in 10'. Where the previous two theories outline how such a response stems from unconscious representational anomalies of the problem, Oaksford and Chater believe the response stems from probability judgements.

This probabilistic view is best exemplified by looking more closely at the rationale behind some of their experiments. Oaksford, Chater and Grainger's (1999) account of the selection task proposes that the material content used in the problem itself can affect card selections. Using the idea of *expected information gain* they have suggested that for the consequent card selections, when the probability of 'Q' is low there will be more 'Q' card selections. When the probability of 'Q' is high, there will be more 'not Q' card selections. By this account, reasoners are seeking rare or unusual information and are therefore maximising their information gain through their choices. The 'not Q' and 'Q' selections are also influenced by the high or low probability of 'P' in an accumulative fashion such that the most 'Q' selections will occur when both the 'P' and 'Q' probabilities are low and the most 'not Q' selections will occur when both probabilities are high. Evans (1998) points out that this theoretical account neglects the influence of 'if' on the antecedent, where there are generally more TA responses than FA (although matching is still taking place).

Oaksford, Chater and Larkin (2000) looked at probabilities, this time using the negations paradigm with the conditional inference task. Here they claim that the probability of the conclusion information is weighted greater than premise information, for all inferences apart from modus ponens inferences, which they claim, are an exception. Thus, depending on the type of logical inference (i.e. MT, DA, AC), the probability of the categorical premise will have a bearing on the probability of the conclusion, depending on the polarity necessitated by the type of inference. For instance, an MP inference with a low probability consequent will have a low probability conclusion (in both cases 'Q'). A DA inference on the other hand will have an affirmative consequent (Q) but a negative conclusion (not 'Q') therefore reversing the probability (or category size).

Probability, negation and category size (contrast class) can be seen as inextricably linked. So much so, that if a small category is negated, the negated clause represents a large category but if a large category is negated, the negated clause now represents a smaller category. That is, negating both smaller and larger categories creates comparatively larger and smaller categories respectively. For instance, a 'collie' (a) is a smaller category than a 'dog' (b). 'Not a collie' (c) is a larger category than 'not a dog' (d) such that a<b<d<c. Negating the smallest category naturally produces a representation of the largest category. The probability of occurrence changes likewise.

Contrast set size and probabilities are very closely related to the mental representation of concept sizes. A large concept could be seen to be one that activates the most related concepts in memory. For instance, the concept of collie is very specific. One might imagine that there are few semantically related concepts in memory e.g. it is a dog, it has four legs, fur etc. The larger concept of 'living thing' on the other hand may have far more semantically related concepts which are also activated e.g. could be a tree, a fox, an amoeba etc and each one of these related concepts also has a whole host of other related concepts. It could be argued therefore, that a larger concept activates a larger area of a semantic network. As the size of the concept increases, so does the probability of something being represented by that concept.

As it is unlikely that reasoners formally calculate the probabilities represented by a named concept, such a calculation could seem to stem from an unconscious or intuitive process. Such a process is solely reliant on previous knowledge – memory. There is a way of immediately representing probabilities just by memory activation alone. If it is considered that on reading an item, it instantaneously activates a pattern in memory of the concept and related concepts. If two concepts are encountered then two patterns of activation may take place in memory. If for instance the concepts of dog and animal are represented in memory by the activation of semantically related networks it is likely that the concept of 'dog' may in some way overlap in its activation with the concept of 'animal'. The degree of overlap in terms of area or spread of activation between the activation of 'dog' and 'animal' could resemble the probability of a dog being an animal. Simplistically, if the concept of 'animal' was represented by the activation of a 100 concepts that are all animals, then the concept of 'dog' may overlap one of these concepts already stored as an animal. The intuitive probability then of a dog being an animal would be 1 in 100 represented just by the degree of activation overlap. Probabilities represented by concepts can therefore be *perceived* in memory rather than calculated. Probabilities of concepts and memory activation areas caused by concept size can not easily be disentangled experimentally.

Summary

It has been shown that responses to conditional reasoning problems are heavily influenced by the content contained within the problem itself. The logical requirements set by the problem do little to influence the way people generally respond to the problems. Even other logical formats using universal, disjunctive and negative conjunctive propositions, seem to elicit the matching bias phenomenon. Thus, the main influencing feature of processing a problem is the initial representational stage. It is at this stage that content has the greatest effect. Using realistic or thematic content can increase logically correct responding and decrease matching bias, perhaps only when the content provides a semantically enriched representation of the problem. It would seem that abstract tasks elicit matching bias because the content is semantically impoverished.

Also, matching bias only occurs in abstract tasks when the negations within the problem are displayed implicitly. This manipulation reduces the responses given only to those inferences that contain implicit negations (mismatching concepts) and thus creates the matching bias effect. It will be further argued in Chapter 3 that an implicitly negated concept no longer activates the same semantic network as the item in the premise. This lack of activation overlap is seen to reduce the perception of relevance of that item with regard to the premise. These changes in problem content can only be changing differences in the way memory represents these concepts. Everything else such as logic, has been held constant.

The three main theories that describe matching bias all posit the influencing nature of the problem content, either in the initial representation of the problem or in the way the content is interpreted. Apart from the optimal data selection account which suggests the influencing nature of the probabilities that can be interpreted from the items mentioned in a problem, there is little mention of the mechanisms that may be responsible for some items appearing more relevant than others. It was argued that the process of memory activation can account for both heightened perceptions of relevance in forming an initial System1 representation of the problem and can also account for apparent probability effects.

The following chapter will look at the general findings and theoretical stances concerning implicit memory. It will be shown that implicit memory and learning are powerful unconscious processes that can affect a person's behaviour in terms of responses to simple word identification tasks. Here, previous exposure to an item at levels below conscious awareness can facilitate or interfere with responses to subsequent tasks. It will be argued that such unconscious influences may be similar to the unconscious perceptions of relevance (System1) that are seen in problem solving. A more in-depth look at some of the current findings in reasoning will take place in Chapter 3 in which comparisons will be made with the research on memory outlined in the next chapter. Chapter 3 will then conclude with a proposed model of reasoning biases based on the memory research findings and subsequent theories of memory processes.

Chapter 2

Memory

Introduction

This chapter will start with an overview of one aspect of memory research – namely implicit memory (and learning). As there are major theoretical viewpoints in reasoning concerning a possible unconscious perception of what is seen or represented as relevant aspects of a problem, it is worth taking an overview of the memory literature in which again unconscious mental processes are the major focus. The phenomenon of 'priming' will be looked at whereby previous exposure to a stimulus affects responding on a later task and it will be proposed that the underlying mechanisms responsible for such an effect may also be responsible for biased responses on reasoning tasks. The chapter will then lead on to local and distributed network models of memory (Collins & Loftus, 1975; Anderson, 1976,1983) looking at some of these earlier theoretical stances, aspects of which have endured to the present day and which provide a good level of explanation of what might be the cause of underlying biases in reasoning tasks. Parallel Distributed Processes (PDP) and connectionist models of memory (e.g. McClelland & Rumelhart 1985,1986) will follow and the focus will turn to how the issues of memory representation and reasoning biases may combine.

Implicit Memory & Learning

The phenomenon of an implicit memory system is noteworthy when discussing unconscious influences of the materials used in reasoning problems. The bias proposed to exist in reasoning stems from an unconscious, intuitive notion of what is seen as relevant, and fits in well with reasoning theories such as Evans' (1984) heuristic-analytic account and Johnson-Laird and Byrne's (1991; 2002) mental models account. It is reasonable to assume that if memory processes were underlying these reasoning biases, then it is not because reasoners are trying to consciously remember the materials. Such influence in a reasoning problem would undoubtedly happen without their conscious awareness.

The notion of an implicit memory - that is, the retrieval of a previously learned stimulus without conscious recollection is relatively quite recent in experimental psychology. Until the seventies it had not been rigidly scrutinised under experimental conditions. In the last two decades there has been somewhat of a glut in the emergence of implicit memory studies. These studies have generally emerged from disparate quarters, whereby the findings supporting the existence of separate explicit/implicit systems have been rather as by-products from studies whose initial main intention was focused in other areas. The overview will start by delineating differences between explicit and implicit memory. Many of the studies of implicit memory have taken place investigating amnesic patients. The phenomenon of interest here is the distinction between explicit and implicit memory systems. Amnesic patients can show impairment on direct explicit memory tasks but similar facilitation on implicit tasks to normal subjects.

Explicit/Implicit Distinction

A main line of investigation using these memory paradigms is that of discovering whether there are any qualitative differences between the explicit and implicit systems. This line of research has been inspired by the criticisms that there are only quantitative differences between an apparent implicit system and an explicit system and therefore could be one and the same. Jacoby and Dallas (1981) carried out a study in which, instead of the standard paradigm involving the presentation of a study list, they required participants to answer questions concerning the target words, which meant they were performing elaborative processing. Explicit memory for the recognition of the words was higher in the elaborative tasks than the non-elaborative tasks. This manipulation however had no effect on the implicit system, with priming effects being similar across both conditions. Graf and Mandler, (1984) showed differences between the explicit and implicit systems *just* by differences in the task instructions. In the explicit conditions, participants were asked to remember the items from the initial study episode given a cue, whereas in the implicit condition, participants were instructed, given the same cue, to think of the first word that comes to mind.

Similarities though, have been observed between explicit and implicit memory using elaborative processing of the study task in relation to memory for new associations. Schacter and Graf (1986) showed impairments in both types of memory when the study task involved elaborating semantic links between two unrelated words. When participants elaborated on these semantic links, memory in both aspects was facilitated. When the study task prevented the elaboration of semantic links by creating an alternative task such as comparing amount of consonants and vowels between the two target words, performance declined in both memory systems.

There have also been findings concerning the nature of the context of the study episode and the context of the testing episode. It has been found for instance, that if the modality changes between the two conditions (from written to auditory), this impairs implicit memory but not explicit. Also, there has been some fragility shown of the implicit system when using the same modality. Groeger (1986) found that when given the sentence, "She looked ______ in her fur coat", followed by two possible completions (which were semantically different); smug and cosy, those that were primed with the word 'snug' as a subthreshold auditory prime were more likely to choose the semantically related word 'cosy' to complete the sentence. Those given the same auditory prime above threshold, but below a level for correct identification, were more likely to complete the gap with the similar sounding phonological option of 'smug'. Thus, there would appear to be a semantic/phonological distinction dependent on whether the prime was delivered at an unconscious or conscious level (see also Groeger, 1988).

It would appear that there may or may not be two separate underlying mechanisms for implicit and explicit memory phenomenon. The implicit/explicit distinction concerns the instructions or requirements demanded by the memory task. No such requirements are asked for in a reasoning problem, therefore any effects of materials can be assumed to involve implicit processes. These biases are seen as implicit in the respect that they affect behaviour at an unconscious level. A reasoning task does not necessarily require an explicit recollection of the problem materials because they are all immediately available when a decision is to be made. Biases that derive from reasoning are believed to stem from unconscious implicit processing of the materials and it will be seen that such unconscious effects also exist in priming tasks. Parallels can be drawn between the underlying implicit processes that are responsible for priming and those that may be responsible for matching items in a reasoning problem being perceived as more relevant.

Priming Phenomenon

The concept of priming basically describes how exposure to a stimulus on a previous occasion can subsequently affect the response to a stimulus on a later occasion. The area of priming research can be broadly split into two types: Repetition priming and semantic priming. It will be seen that there may be components of both types of these phenomena that could cause some choices (mainly matching ones) in a reasoning problem to appear more relevant. There are many issues surrounding different priming effects and these concern the proposed mechanisms that underlie them. It will be seen that different aspects of the tasks may account for repetition priming and for semantic priming, but it is not the scope of the present study to go into great detail concerning all such debates (although the issues will be briefly addressed). More general proposed mechanisms for the occurrence of priming, that link the phenomenon to theories of memory organisation, will be considered later in the chapter. The general point to get across in this section is the fact that pre-exposure to a stimulus can affect a person's subsequent behaviour without their conscious knowledge of it.

Repetition Priming

The area which has predominately spurred research into implicit memory is the phenomenon of repetition priming. It can be defined as the, "...facilitation in the processing of a stimulus as a function of a recent encounter with the same stimulus", (Schacter, 1987), but without any explicit reference to the original study episode. Repetition priming has been considered to be an implicit memory task because the facilitatory effects of prior exposure to a stimulus occur without any *reference* to the initial exposure event. It would be appropriate at this point to list the most widely used paradigms in this line of research. The study or priming phases are relatively similar (although viewing durations change), involving an initial viewing stage of a stimulus, but they differ mainly by how they test for implicit memory. The paradigms that have been used are as follows, with the initial three being the most used: Wordstem/fragment completion (Graf, Mandler & Haden, 1982; Tulving, Schacter & Stark, 1982): Word/perception identification (Feustel, Schiffrin & Salasoo, 1983; Jacoby & Dallas, 1981): Lexical decision (Forbach, Stanners & Hochhaus., 1974; Scarborough, Gerard & Cortese, 1979): Reading of transformed scripts (Masson, 1984; Kolers, 1975, 1976): Free association (Williamson, Johnson & Erikson, 1965): Face identification (Bruce & Valentine, 1985; Young, McWeeny, Hay & Ellis, 1986): Category production (similar to free recall but cued with category labels): answering general knowledge questions (e.g. What animal helped

Hannibal cross the Alps in his attack on Rome?) (both cited in Roediger & Srinivas, 1993).

Typically, participants are given long lists of words to study (e.g. animal), then after a period of filler tasks or even a time delay, they are asked to complete a task usually with the first word that comes to mind (no explicit reference is made to the study episode). In the word completion task for instance, they would be given an incomplete letter stem (e.g. ani_ _ _) or fragment (e.g. a_i_a_). In the word identification task, participants are briefly exposed to a stimulus below the threshold of identification but not perceptual awareness (30 msec) and then they would have to try and identify it. Lexical decision tasks involve being shown a letter string for a short duration as in the recognition task and then being asked a question concerning the letter string, discriminating whether it formed an actual word or not. The other tasks mentioned are fairly self-explanatory and to avoid repetition it is suffice to say that they mainly differ in content from the tasks just described.

With regard to repetition priming studies, there has stemmed two main types of viewpoint concerning what aspect of the tasks may be responsible for the effects, which could be seen to stem from Tulving's (1972) distinction between semantic and episodic memory forms; *abstractionist* accounts and *episodic* accounts. The *abstractionist* position such as the logogen model of Morton (1979) suggests that representations of words are represented in memory by abstract codes. According to Bower (2000), the abstractionist position explains both word identification processes at a perceptual level and offers an indication to short-term priming effects. The abstractionist position for instance, offers an explanation as to why words that are written in different fonts and cases, can still be identified as the same word because all the letters used still map on to the same orthographic code stored in memory. When a word is encountered, the particular code or logogen that represents the word in an abstract way is activated. Priming is seen to be produced by trace memory activations caused by the initial exposure to a word, which then facilitates subsequent activation caused by a repeated encounter of the same word. In Morton's (1979) logogen model, it is suggested that repeated activation of the same abstract representational code lowers a word's threshold so that less activation is required subsequently to identify it. The important point to remember with the abstractionist approach is that it is the activation caused by the perception of the word itself that causes the priming effect.

In contrast, the *episodic* account such as that prescribed by Tenpenny (1995) proposes that it is the event or experience itself that is encoded. A stored record of the encounter is set up each time the word is experienced. Each encounter sets up a new record, and over many exposures all the records some how set up a prototype of that word. Transfer Appropriate Processing (Roediger, 1990) describes how priming occurs because the two similar events overlap in the cognitive operations that stem from encountering the word. The episodic view is basically distinct from the abstractionist view in that it is the unique event and not the actual word (or abstract code) that is being activated.

In attempts to delineate the two approaches, experiments using materials of a semantic nature have been incorporated into the repetition priming tasks in the form of *paired associates*. In effect, these types of experiments combine both perceptual priming, such as those types of study tasks listed above in which the identification of the word is the goal, and also conceptual priming in that access to semantic knowledge is required. By combining these processes needed for the task, insight can be obtained into whether the mere viewing of a word causes priming by accessing some abstract orthographic code, or whether the actual deeper processing of the event causes the priming phenomenon by accessing episodic knowledge.

For instance, Vriezen, Moscovitch and Bellos (1995) used a study in which participants were given a prime and target word (same word) as in the above studies but the task differed in that subjects were required to process the words in the form of a *classification task*. In other words, participants were being cued by the task format to process the words at a deeper more semantic level. In their Experiment 1, they used two semantic classification tasks: is the item 'man-made' and is the item 'larger than a bread-box'. They found that having been given a word such as 'tree', participants responses were faster if they had to make the same semantic classification to the word than if they had to make a different classification. So, if they had to *initially* decide if a tree was a man-made object, on the second viewing of the word they were subsequently quicker to decide again if it were a man-made object than they were to make the alternative classification of a tree being larger than a bread-box. Therefore, they concluded that the semantic nature of the word was being primed (by the classification requirement) and therefore this priming effect carried over to a repeated task in which the processing event was repeated. If it were just the perceptual identification of the word that caused priming effects, then priming effects would have occurred just by the repeated exposure of the word irrespective of the semantic classification episode.

Dennis and Schmidt (In press) used a line of studies in which semantic *comparisons* had to be used. Participants in these studies were given word pairs (e.g. *prosperous/wealthy*) and were required to decide if the words were synonyms or not. They found that if the word *pairing* was repeated, responses were faster than if a subsequent synonym word pairing (made from words that had previously been encountered in the task) was given. For example, if two synonym pairs had been previously encountered (prosperous/wealthy & rich/affluent), participants were subsequently quicker to decide if *rich/affluent* was a synonym than they were to decide about the pair *rich/wealthy*. It would appear that the *association* between the words (episodic event) had been partly responsible for the priming effect and not just the *perceptual* encounter of each individual word.

These studies seem to indicate that whatever is activated by viewing a word it would appear to access not only the perception of the word itself but also its meaning and the circumstances in which it was encountered. It would seem that all these various activations can in some way create a priming effect. Even different task demands can create different amounts of priming. So far, the priming effects discussed have concerned repetition priming where the exact instance or word is repeated. Other priming studies have looked at effects of conceptually similar words to see directly if just semantic processing can cause similar facilitatory effects. The studies that follow indicate that not only are words processed very quickly at both an orthographic and semantic level, but that trace semantic activations of one concept may facilitate activations of similar concepts.

Semantic Priming

McNamara (1992, 1994) in a series of experiments, reports how associated or semantically related words can also cause the priming effect. In one study, a lexical decision task was used in which a response had to be made as to whether a letter-string was a word or not. Using word quadruplets (e.g. mane-lion-tiger-stripes), McNamara showed that a three-step priming effect could occur between apparently unrelated words (e.g. *mane* and *stripes*) when mediated by a chain of semantically related pairs (e.g. lion-tiger). The priming effect was generally weaker than had two directly related words (i.e. shorter semantic distance) been used for the prime and target, but the results offered support for the notion that exposure to words can almost instantaneously activate networks of semantically related concepts.

Typically, semantic priming effects have not been seen to be as durable as the effects seen in repetition priming. Repetition priming effects have been seen to last for hours or even weeks (Bentin & Feldman, 1990), whereas semantic priming often only lasts as long as the experimental trial. Becker, Moscovitch, Behrmann and Joordens (1997) have on the other hand, found evidence for long-term semantic priming. They believe that previous semantic priming experiments have failed to show long-term effects because the level of processing of the items (prime & target) was not deep or elaborative enough and the degree of semantic overlap between the items was not great enough.

Becker et al (1997) used a semantic decision task in which participants had to decide if words (concepts) were 'living' or 'nonliving'. This was compared to a condition in which a lexical decision had to be made i.e. is the item a word or not. The experiments differed from previous ones in that multiple semantic primes were used before the target word was encountered. This enabled repeated activation of similar semantic networks. Also, the semantic decision of living/nonliving required a greater depth of processing to take place. The findings showed semantic priming effects lasting more than two minutes, which were not found on the lexical-decision task.

Becker et al (1997) suggest that priming is a form of 'incremental learning' which is accountable for all forms of priming. This incremental learning hypothesis describes in terms of connectionist architectures, how connections that are activated upon encountering a concept are in some way altered to facilitate further activation of that concept. Before considering the theoretical aspects in relation to reasoning biases, it is necessary to give further details concerning connectionist models of memory. The following section therefore gives a brief outline of some of the issues raised so far before looking at distributed models of memory and connectionist systems and how possible priming effects may be responsible for reasoning biases.

Models of Memory and Priming Accounts

It is far beyond the scope of this chapter to review the whole area in detail (see Chang, 1986; Schacter, 1987; Richardson-Klavehn & Bjork, 1988; Joordens & Becker, 1997, for reviews). There are a whole host of categories that various studies, theories and paradigms have been placed in by various authors. Schacter's (1987) categorisation of memory models will be used here for theoretical interest and completeness but only a brief outline will be given of each. It is felt that these compliment the later section on network models of memory whilst not overlapping in any considerable way. As previously pointed out, two of the memory models concern the idea that there are separate memory systems or processes using either Tulving's (1972; 1983) episodic/semantic distinction, declarative/procedural systems or differences in encoding and retrieval systems. Whilst semantic and declarative systems share the same notion that there is a separate system for storing 'data', the episodic system describes temporal and personal memories and the procedural system describes the knowledge underlying skilful behaviour. Schacter's three types of models are categorised as; *activation*, *processing*, and *multiple memory* systems. Richardson-Klavehn and Bjork (1988) categorise the positions into *abstractionist* and *nonabstractionist* philosophies. The main difference between views here is whether the 'activation' models are separate or part of an abstractionist (multiple memory) position.

The activation account states that priming is caused by the activation of existing knowledge representations and it is this temporary activation that causes subsequent facilitation in the tasks. It is tantamount to the trace activations that will be discussed in the section on network accounts although here as stated by Schacter (1987), there is still the mention of a separate elaborative mechanism associated with episodic memory traces.

Supporting evidence for the activation view stems from the finding that priming of already acquired knowledge/representations operates independently from elaborative processing and therefore works at a very basic level (Graf, Mandler & Haden, 1982; Jacoby & Dallas, 1981). The activation view does not account for the temporal durability that has been found to last days and weeks on certain, previously mentioned priming studies nor does it account for recent findings concerning the priming of unfamiliar non-words (Stark & McClelland, 2000).

The processing account (see also, Richardson-Klavehn & Bjork, 1988 for a review of non-abstractionist positions) derives from initial work instigated by Kolers (1975; 1976). It's postulated that viewing a stimulus such as a word requires a set of sensory-perceptual and conceptual analysing operations, and that when these operations are instigated it constitutes a form of practice and skill acquisition. Therefore making any subsequent operations concerning that stimulus easier when encountered again. A further processing approach in this category is the idea of *transfer appropriate processing* (TAP - Morris, Bransford & Franks, 1977) which proposes that priming is facilitated by overlap between the processing of the study item and the processing of the test item. It proposes that the greater the similarity, the greater the overlap in the processing requirements. Although it has been pointed out (Graf & Masson, 1993) that this account seems to be consistent with any experimental outcome and cannot explain differences observed between explicit and implicit memory test performance, which is why TAP is now only used as a general framework.

Roediger and Srinivas (1993) concentrate on the distinction between perceptual and conceptual processing. They still maintain the notion of processing overlaps but assert that the implicit affects are caused by perceptual processing overlaps and the explicit affects by conceptual processing overlaps. This follows on from earlier views (Jacoby, 1983) that conceptually driven processes reflect subject-initiated activities such as elaborating, organising and restructuring, whereas perceptual (data driven) processes are guided by the immediate information offered by the stimulus. Other processing accounts have been; integrative vs elaborative processing, interpretive vs elaborative, and environmentally driven vs subject guided processing. However, Schacter (1987) points out that the processing account has difficulty encompassing the short lived effects of some priming techniques and the dependence in certain cases of amnesic patients, of priming of pre-existing knowledge. Also, in the stem completion task, priming for new associations can depend on the amount of elaborative processing done at the study stage which is not easily accounted for by the processing approaches.

Finally, there are the independent memory system accounts. These generally postulate that there are different memory systems associated with different brain structures (abstractionist positions; Richardson-Klavehn & Bjork, 1988). Squire (1986) proposed two kinds; procedural and declarative. It is the procedural system that is thought to survive in amnesic patients. The differences between the two systems can best be described as 'knowing how' and 'knowing that'. 'Knowing how' involves perceptual and motor operations, and 'knowing that' concerns knowledge about the particular time and place an event has happened. Tulving (1972; 1983) also proposed a distinction in memory systems between episodic and semantic memory, representing explicit and implicit findings respectively. Again, episodic memory is used for events, whilst semantic memory is used for existing knowledge. These accounts however, fail to describe the findings observed in amnesic and normal participants. According to Roediger and Srinivas (1993, pg20), "...most researchers working within this framework would agree that at least half a dozen memory systems are needed to explain the dissociation between tests and the number may extend to about 25".

Local and Distributed Network Models

The common theme between both of these approaches is that memory is highly organised by way of semantic similarity. Activation of one concept will lead to activation of a similar concept. Differences between the two may lie more in their levels of explanation than their philosophy. Local networks offer a schematic conceptual view of memory whereas the distributed, neural network approaches are far more elaborate in their processing descriptions. The former approach is viewed as a serial, sequential process whereas the latter is seen as a more powerful parallel process. Concepts in the local model though are seen as being abstract representations as a whole i.e. the concept of 'cat' is seen as having its own specific neural location – a node is activated that represents 'cat'. In a distributed model, a concept is represented by the action of several nodes that represent microfeatures (Hinton, McClelland & Rumelhart, 1986). A cat might therefore be represented by the *pattern of activation* of several nodes for the concept of cat. There may be a thousand nodes activated just for prototypical feature representations for instance. There are aspects of both network approaches that are relevant to this study.

Local Networks

Collins and Quillian (1969; 1970) initially produced the hierarchical network model of semantic memory, which later was revised into the spreading activation theory (Collins & Loftus, 1975). In Collins and Quillian's (1969) hierarchical model, they propose semantic memory is organised in superordinate hierarchies so that a concept such as 'canary' would come under the category of 'bird' which in turn would fall under the superordinate category of 'animal'. At each point in the hierarchy (superset), certain properties would be stored against each item. That is, next to 'canary' for instance, would be stored the properties 'can sing'/'is yellow'. The theoretical assumption here being that the further away two pieces of information in the hierarchy are (property or superset), the more time it would take to make decisions about them in a verification task for example. So that, 'a canary can sing' would be quicker to verify than 'a canary has skin', - with singing being a direct property of canary whilst 'skin' being a property of 'animal'. Therefore, to ascertain whether a canary has skin would involve the further inference that a canary is a bird and a bird is an animal.

It is interesting at this point to see that within their hierarchy, Collins and Quillian also include negated instances e.g. 'an ostrich can't fly'. This is seen as an important feature of an ostrich and by including the negation allows the general feature 'can fly' to be included under the superordinate category's ('bird') *general* features. All the general prototypical information only needs to be stored once for 'bird', and the exceptions can be added to individual cases. Any search of memory for validation/contradiction purposes thus scans the list of properties at each node, then moves up to scan the properties at each superset. What is unclear is at what point the scanning process comes to a halt for 'false' (negated) responses i.e. when does the search stop when given the question 'is a canary blue'. Several possibilities have been offered. One possibility is that the search stops when encountering a refuting property (canary is yellow) although their data do not support this notion. They have also suggested that memory is searched for a limited amount of time, although intuitively one would imagine that the search really depends on the task at hand. You could spend all day trying to retrieve the name of a familiar person you saw. Thirdly, they offered the idea that memory was searched for a limited number of levels or 'depth' within the hierarchy. All these explanations come under the scrutiny of intuitive counter-examples and the whole notion of the links between nodes, supersets and properties becomes less predictive.

The spreading activation theory (Collins and Loftus, 1975) is founded within the realms of artificial intelligence. The revised model incorporates a more elaborate system of links between the nodes or concepts arranged not so much within a hierarchical organisation, but a distribution of clusters representing interrelated concepts. The concepts of 'vehicle', 'truck' and 'fire engine' being clustered together whereas 'colours' form another cluster. To recall that a fire engine is red, a link would have been formed from the one concept to the other. The more features two concepts have in common, the more they will share the same links. This model has been criticised by Chang (1986) for being too general, difficult to investigate empirically and low on predictive power. The model though, may well have been the forerunner to the more recent distributed approaches to memory research.

Along similar lines, although offering more refinement is Anderson's (1976;1983) theory of Adaptive Control of Thought (ACT*). Here, the basic tenets of knowledge are encoded into the primary propositions that the knowledge espouses. So instead of nodes representing concepts as in the spreading activation model, here the nodes represent simple propositions (e.g. the dog chased the cat). Using the example, the node representing the proposition will be linked to the concepts of 'dog' and 'cat'. The link between the concepts and propositional node is represented by the relationship between them. So 'dog' will be linked to the propositional node by the relationship link of say, *agent*. 'Cat' will be linked to the node by the relationship *object*, and 'chase' will be linked to the propositional node by *relation*, with the episodic concept 'past' being linked by *time*. The network is therefore more richly represented than the spreading activation hierarchy previously mentioned.

The links between nodes are not necessarily equal in the ACT* model, with some links being more associated with nodes than others. This would cause the greater associated links to create more activity in those related nodes than other nodes whose links were weaker. This facilitatory effect explains typicality effects (as in the form of concept prototypes) as well as familiarity ones (multiple exposure) in that previous repeated activations can account for stronger, more associated links. There are also limits to the total activation that can take place between nodes. This total amount of activation appears inflexible in that if only one node is activated with a few corresponding links, then these will be activated more than if there were many nodes with links of near equivalent association (which would dilute the total amount of activation). There would appear to be a fixed limit of total activation that takes place and it is neither increased nor decreased – just distributed.

Anderson claims the spread of activation accounts for a *fan effect*. If a concept activates a large amount of equally associated nodes, then activation of all these nodes will take longer than if just a few related concepts were activated. The spread of activation of related nodes is indiscriminate, with the only guide to their activation being by their relatedness. The fan effect would therefore predict that a more complex associative network of concepts and propositions would be slower to be activated. If we take the category size of a concept as a possible indicator of the amount of stored representations connected to that concept, then a concept representing a large category like 'living thing', may be slower and more difficult to process than an instance of that category such as 'snake'. It might also be argued that activation of the 'snake' concept may be greater in some way because the total amount of activation power is only being used to activate a few nodes. Activation of these nodes will also happen more quickly, being fewer in amount to the concept 'living thing'.

Activation amongst nodes and links spreads and fades out (presumably as it becomes more 'diluted'), but there is scope within the theory of a rehearsal mechanism which can maintain the activation of certain nodes. There is also a direct two-way link here with attention in that what is attended to is activated, but if something is activated to a high enough level (compounded activation) it will reach conscious awareness. A concept may have some trace activation due to being *related* to what is being directly attended to – not necessarily being attended to itself. If this trace activation is increased (possibly due to the activation of another similar stimulus), this compounded activation could then reach conscious thought (it has already been seen in the priming studies that activation can occur without any conscious awareness of the stimulus or concept). The implication is that there is an upper level of activation at which conscious thought begins and in effect, two or more 'unconscious' thoughts can in fact produce a conscious thought as long as the activations overlap.

Parallel Distributed Processing/Connectionist Accounts.

Rumelhart, Hinton and McClelland (1986) offer a theory of mental processing based on parallel distributed processing (PDP). Here, high level concepts such as the instances used in the categorisation tasks such as dog, animal etc. are not represented in memory by distinct 'nodes', but are instead represented by a pattern of 'unit' activations – where each unit represents a minor feature of the stimulus ('microfeatures' – Hinton, McClelland & Rumelhart, 1986). Basically, where a stimulus is encountered, a pattern of activation occurs, in which all the units that represent features of the stimulus are activated. Close approximations to this pattern or activation overlaps inspired by activation of several similar concepts come to represent the prototype of the stimulus category.

A set of units can only contain one pattern at a time, which in turn creates an 'enforced seriality' of processing. Rumelhart et al (1986) point out that any processing which only takes milliseconds is probably a parallel process, whereas any processing that takes a few seconds is a series of parallel processes and is therefore sequential. We might find therefore, that experimental paradigms which induce very short reaction times such as the sentence verification task and those looking into semantic priming effects may be looking directly at the way memory is organised. Whereas other paradigms such as those used in reasoning may be susceptible to both the initial parallel activation caused by the stimulus (System1) in addition to subsequent serial activations caused by trying to 'work out' the problem at hand (System2). Thought processes themselves are said to be constantly driven by external stimuli. This 'data-driven' view does not preclude the notion that internal drives can also cause activations. There is said to exist an internal system of 'mental models' that are driven by hypothesised actions (mental simulations) or driven by 'recycling' linguistic inputs. It would seem that internal drives might be in place with respect to a kind of System2 process, whereas more direct (faster) activation occurs from processing external sensory stimuli and would be akin to a System1 process. It might be that the representation of a problem, together with internally driven representations to solve the problem, may in some way overlap or interfere, thus producing a seemingly 'intuitive' response in solving the problem but actually being biased by the materials used and their subsequent mental patterns of activations.

Rumelhart, Smolensky, McClelland and Hinton (1986) claim that problems are solved not so much by the use of logic, but by, " making the problems we wish to solve conform to problems we are good at solving". The implication being that we re-organise the structure of a problem and map it onto problems we have encountered before. Even though they state that formal logic is rarely if at all used, they outline three abilities which enable people to come to logical conclusions: Pattern matching – we can easily and quickly represent internally an external stimulus: Modelling – we can anticipate external states of affairs by creating prospective models: Manipulating the environment in order to make sense of it. Generally they claim that answers to problems are 'perceived'. On expert tasks such as chess playing, complex move alternatives are reduced to pattern matching – the patterns having been created by previous experience. This is essentially tantamount to an implicit, powerful system that delivers 'intuitions' to problems, and is to a large part dependent on the way the problem is cognitively represented. It could be that with problems that look *superficially* straightforward as in conditional reasoning tasks, people seek internally, a fast intuitive answer, possibly based on trace activations of unit patterns.

Rumelhart, Smolensky, McClelland and Hinton (1986) argue that schemas are represented by coalitions of activations of units and it is this 'pattern of activation' that comes to broadly represent a schema. Slight deviations from these patterns e.g. seeing a hand-written number '4', still activates enough units to form a pattern (or coalition) which maintains a ''goodness of fit'' with that which represents the '4' schema. If schemas are 'rigid' i.e. a '4' can not be represented by a 'Z', then they suggest the peak of activation must be very narrow in the 'goodness landscape'. If on the other hand, the schema is not so rigid i.e. a plant, then the goodness landscape would contain broader peaks (or plateaux). Therefore, memory and subsequent recognition, would be a case of matching the goodness of fit of the peak caused by the external (or internal) stimulus, to the internal memory representation of the peak of activation for that schema.

McClelland, Hinton and Rumelhart (1986) outline the use of units that can represent or correspond to minute aspects of a stimulus e.g. the top left-hand corner of a cube. It is appreciated that what is actually represented by the units at this level is unknown. It is postulated that these units are highly interconnected with each other. A mass of these units is classed as a 'module' and modules themselves are highly interconnected. The connections between units carry 'weights' or strengths which can change as the patterns of activation between units come and go and the whole activation sequence reaches a kind of equilibrium and comes to 'rest'. As described previously, input into a module can come from external stimuli and/or from input from other modules. So any stimuli, either externally or internally, creates a pattern of activation amongst units within modules and across modules. A visual stimulus such as the word 'animal' may initiate a large pattern of activation of all the related units associated with viewing the word. Recognition of the word (or memory of it) stems from a series of changed weights on the connections between units. Such changes enable the pattern of activation to reach a resting point sooner than it might have, had the stimulus been novel. McClelland and Rumelhart (1986) make no attempt at describing the temporal duration of the changed weights, although they do indicate that the change in weights may be more robust after repeated similar patterns of activation. The traces decay rapidly at first, but the remaining portion becomes more resistant through repeated activations. They do not exclude the notion that weights can change after only one exposure, but the implication is that this change may not be long-lasting. Unit activation is said to decay much faster than changes in weights.

Priming and Reasoning Biases

It is quite obvious that the priming tasks are far removed in their procedures from conditional reasoning tasks. The conditional reasoning tasks display all the concepts at the same time rather than at separate occasions (as is the case for target and prime words in priming tasks). The duration of the exposure to the stimuli in reasoning tasks is within the control of the participant – they can view the materials for as long as they like until they make a response, whereas the onsets are strictly controlled for in typical priming studies. Many of the priming studies (especially repetition priming) concern an implicit memory system in that conscious recollection of the initial study episode can not be made. Whereas in the reasoning tasks, recollection does not have to be made at all as all the materials are immediately available to conscious thought processes.

The differences between the two types of task may be more apparent than real though, especially if looked at in terms of the possible underlying processes that may be involved. It would not seem to be a prerequisite for priming to take place, in that the materials have to be displayed on different occasions – this is more an incidental occurrence of the way the priming paradigm has evolved. None of the explanations that have been offered for priming suggest that separate encounters of the prime and target words are fundamental to the effect taking place. There is no suggestion that seeing both items together at the same time will detrimentally affect the priming of one item to the other.

As far as the duration of the exposure to those items, the study by Becker, Moscovitch, Behrmann and Joordens (1997) suggests that the more exposure to semantically related items, the more priming that will take place. Also, they suggest that the greater the level of semantic processing, the larger the priming effect. The semantic priming studies also show that it is not necessary whether participants can consciously remember the initial exposure to the target item or not. Where the implicit/explicit distinction seems to lie is in respect to the fact that participants are not consciously aware of the priming effect. The *effect* would appear to be unconscious, rather than the conscious awareness of being exposed to the stimuli.

It is the contention of the present study that the underlying processes responsible for priming are also responsible for the matching bias effect seen in both the Wason selection task and the conditional inference task. It is argued that the conditions exist in a reasoning problem that are ideal in order to facilitate priming. Bowers (2000) points out that although many words are encountered on a daily basis, not all these will produce priming effects because they are encountered only for brief periods. During the normal reading of a text, there is an average fixation per word of only 200-250 msec. This he believes, is too short a period for any relevant changes to take place. The implication then, is that the longer the exposure to a word (and also the greater the depth of semantic processing that is being carried out), the greater the priming effect that will occur. Therefore, the standard reasoning tasks in which participants (probably) fixate on the items in the given premise for long periods and presumably consciously process the material at deeper levels, are ideal situations for priming to take place.

Priming effects can be explained by the distributed model in terms of trace activation. What has been input previously (e.g. external visual stimuli) causes a pattern of activation which is subject to trace decay. On subsequent viewing of the same (or similar) stimuli, the pattern of activation is facilitated by the trace decay of the change in weights. Such facilitation could contribute to a reduction in response times to previously experienced stimuli. With respect to semantic priming, it is seen that the subsequent exposure need not be exactly the same as the previous one. As long as a 'similar' pattern of activation is initiated, slight changes in subsequent input would in some way 'overlap' the initial 'composite'. The overlap between patterns could represent prototypical features of the two inputs which form the concept that is being described, either as an individual item (e.g. 'dog') or a categorical representation (e.g. 'a living thing'). Extrapolating this notion further, one could also explain the semantic priming of one word/item to another by viewing the process in terms of pattern overlap and residual trace decay.

One might presume that such semantic overlap may not be as great as the overlap caused by viewing the same word. In which case it might explain why semantic priming effects are generally not as durable as the effects achieved with repetition priming. It is worth noting that McClelland and Rumelhart (1986) claim that traces can coexist within the same memory domain. That is, more than one trace can exist at a time even if it involves some of the same units. It could be that these trace activation patterns interfere in reasoning by causing perceptions of some of the concepts mentioned in the problem as appearing more relevant than others. This could lead to unconscious intuitions concerning the appropriate answer for the problem. In the next chapter, the issue of these trace activations interfering with the task will be redressed with an attempted assimilation with the reasoning literature.

Chapter 3

Reasoning Biases & Memory Activation

This chapter sets out to explain matching bias and difficulty with processing negations in terms of distributed models of memory and will attempt to make clear and precise predictions concerning the content of conditional reasoning problems. The main reasoning biases, and perhaps the only ones, are *matching bias* and *double negation* effects. Matching bias was outlined in Chapter 1 and will be recapped in more detail here together with details on the double-negation effect. The present chapter will start with an outline of the terminology and the ways in which aspects of the 16 negations paradigm inferences have been looked at. Due to the structure of the four inference types that are usually studied (i.e. modus ponens 'MP', modus tollens 'MT', denial of the antecedent 'DA' and affirmation of the consequent 'AC'), any systematic changes of negation in the major premise also alter the negations in the minor premise and conclusion. Several terminologies have been used by different researchers in order to divide the 16 conditional inference types up in different ways. It will be pointed out that these subdivisions may be measuring the same phenomenon and that the terminologies used may be misleading.

If the 16 negations paradigm inferences are looked at solely by the way the concepts that refer to each other in the problem either match, partially match or do not match, a clearer picture emerges concerning content effects on the frequency of inference endorsements, and hence biased response patterns. When the inferences are looked at in this way, a simple model of unconscious distributed memory activation can account for many, if not all of the biases that have been reported using abstract reasoning tasks. This model will be described in detail near the end of the chapter after looking into the details of negation effects in both reasoning and memory studies. Negation is seen as instrumental to the presence or absence of matching bias and obviously double-negation effects.

Response Patterns Caused by Negation

Before discussing negation and reasoning, a few points with regard to terminology need to be outlined. It will be remembered that in the first chapter, the negations paradigm was discussed in relation to standard conditional problem formats. Using the conditional inference as an example, there are deemed to be two correct forms of response (MP & MT) and two typically fallacious responses (AC & DA). There are three aspects to a conditional inference problem: major premise, minor premise and conclusion. E.g.

<u>(Modus Ponens)</u>

(*major premise*)	If there is a cat, then it has four legs.
(*minor premise*)	There is a cat.
(*conclusion*)	Therefore it has four legs.

These can be broken down into their constituent parts by the way they refer to each other. Using the above example, the concept mentioned in the minor premise ('a cat') is referred to as the *categorical premise.* The categorical premise in this instance refers to the concept 'a cat' in the major premise, which has been labelled by Schroyens, Verschueren, Schaeken and d'Ydewalle (2000) as the *referred clause*. The *conclusion* ('has four legs') refers to the major premise clause 'has four legs' which is termed by Schroyens et al (2000) as the *inferential clause*. By systematically changing the polarities (the presence or absence of 'not') of the items mentioned in the major premise we obtain the four inferences of the negations paradigm (from Chapter 1, *Negations Paradigm*):

If A then 2

If A then NOT 2

If NOT A then 2

If NOT A then NOT 2

By applying the four negations paradigm premises to all four inference types (MP, MT, AC, DA – see Chapter 1, *Conditional Inference Task*) produces 16 conditional inferences (if only implicit negation is used in the categorical premise – see Appendix A for a full outline).

Affirmative Premise Bias

These 16 conditional inference types have been categorised in the literature by several methods depending on the focus of research interest. For instance, Evans, Clibbens and Rood (1995) looked at the frequency of inference endorsements along the dimensions of *affirmative premise bias* and *negative conclusion bias*. Affirmative premise bias could be seen as a tendency to endorse more conditional inferences that contain affirmative minor premises than negative ones. For instance, using the two modus ponens inferences:

(1) If A then 2	*(2) If NOT A then 2*
A	*NOT A*
Therefore 2	*Therefore 2*

It can be seen that (1) has an affirmative minor premise and (2) has an explicitly negated minor premise. Affirmative minor premise bias would be seen by more endorsements to (1) than to (2).

Such a biased response pattern had been largely unfounded in the literature but it was Evans' (1993) contention that the model theory of conditionals should predict this bias. Evans and Handley (1999), found for the first time that when *implicit* negation (i.e. 'B' instead of 'not A') is introduced into the minor premise in a conditional inference task, reasoners were more likely to accept an inference consisting of an affirmative minor premise than a negated one. Previous to this study (e.g. Evans, Clibbens & Rood, 1995), conditional inferences had been expressed solely with the use of explicit negation. Using implicit negations creates the comparison between inferences which *match* (affirmative minor premise) and which *mismatch* (negative minor premise). The affirmative premise bias (using implicitly negated minor premises) is thus equivalent to a measure of matching bias.

Negative Conclusion Bias

Negative conclusion bias would be seen when there is a tendency to endorse more inferences that contain negations in the conclusion. For example, using the two modus ponens inferences:

(1)	(2)
If A then 2	*If A then NOT 2*
A	*A*
Therefore 2	*Therefore NOT 2*

Negative conclusion bias would be seen if there were more endorsements to inferences with negative conclusions (2) than affirmative ones (1). There has tended to be weak overall effects of negative conclusion bias for *modus ponens* (MP) and *affirmation of the consequent* (AC) inferences. The effect is far more pronounced for denial inferences (MT & DA) which led Evans, Clibbens and Rood (1995) to conclude that reasoners were having difficulty with *double-negation.*

Double-Negation Effect

The double negation effect occurs on inferences that require the falsification of an already negated inferential clause. For instance;

(Modus Tollens)

If the letter is NOT an A then the number is NOT a 2.

The number is a 2

Therefore, the letter is an A

Reasoners see that the categorical premise ('2') refutes the referred clause ('NOT 2') and then an attempt is made at refuting the inferential clause ('NOT A'). Thus to reach the conclusion ('A'), reasoners must make the inference 'NOT(NOT A)'. Difficulty with processing this double negation would appear to be responsible for the finding that reasoners will be less likely to endorse such a conditional. So it is not that reasoners are more likely to endorse negative conclusions, but instead are *less* likely to endorse affirmative ones on MT and DA inferences.

Alternative Terminology

As previously pointed out, there would seem to be two main biasing phenomena in conditional reasoning; matching bias and double negation effects. Other commentators though may have inadvertently re-labelled these by re-categorising the 16 negations paradigm inferences. An example of this can be seen in Schroyens, Schaeken and d'Ydewalle (2001). In their meta-analysis of several conditional inference experiments, they have classified the 16 inferences into two categories; *inferential negation* and *referred negation*. Both concern the presence or absence of negations in the major premise on both the antecedent and consequent clause. That is, when the referred clause is negated (referred negation) or when the inferential clause is negated (inferential negation).

Their findings concerning *inferential negation effects* reproduce Evans et al's (1995) findings of a double negation effect in that fewer inferences were endorsed in both DA and MT inferences when the inferential clause was negative. In DA and MT inferences when the inferential clause is negative, the inference requires the denial of an explicit negation i.e. not 'not A'. As the inferential negation effect was mainly limited to DA and MT inferences, their results have re-interpreted the double-negation effect.

On *referred-negation effects*, they found that reasoners are more likely to endorse an AC inference with an *affirmative* referred clause and they are more likely to endorse a DA inference with a *negative* referred clause. This is essentially matching bias. In both AC (affirmative referred clause) and DA (negative referred clause) instances that produce a larger frequency of inference endorsements, the inferences also have affirmative categorical premises. The following AC and DA inferences are used as examples:

(1) AC	(2) DA
If NOT A then 2	*If NOT A then 2*
2	*A*
Therefore NOT A	*Therefore NOT 2*

In both cases the minor premise item matches the referred clause item. The referred clause in (1) is affirmative and in (2) negative. It would seem that matching bias on the conditional inference task (using implicit negations) has acquired the labels of *affirmative premise bias* and *referred negation effect.*

It has been highlighted that the variety of ways in which the 16 negations paradigm inferences can be divided up creates a series of *apparently* different effects. These different effects may be more apparent than real, in that they may cross over with already established effects such as matching bias. While it seems apparent that matching bias and difficulty with negations are genuine phenomena, caution must be maintained through restructuring analyses and re-labelling. A clear direction through all this is to look specifically at the problem structure and the polarity of the materials that are used. In response to Evans et al's finding concerning the role of matching bias and implicit negation seen in Chapter 1, Schroyens, Verschueren, Schaeken and d'Ydewalle (2000) have taken a closer look at the relationship between the categorical premise and the referred clause. They make the distinction between *explicit/implicit affirmation/denial* inferences. This distinction sheds more light on the actual role of implicit/explicit negation and matching bias.

Explicit & Implicit Negation

It will be recalled that an explicit negation contains the original referent preceded by the word 'not'. An implicit negation is given as an example of an item that is not the original referent. So that an implicit negation of the concept 'not a cat' could be 'dog' whereas this can be stated explicitly as 'not a cat'. For instance, in two conditional inference problems of the same logical structure, the minor premises differ in that;

<u>(Explicit negation – Modus Ponens)</u>	<u>(Implicit negation – Modus Ponens)</u>
If the letter is NOT an A then the number is a 2.	*If the letter is NOT an A then the number is a 2.*
The letter is NOT an A.	*The letter is a B.*
Therefore, the number is a 2.	*Therefore, the number is a 2.*

Schroyens et al (2000) point out that the explicit/implicit manipulation only changes half of the 16 negations paradigm inferences. This is best highlighted by Table 3.1.

<u>Table 3.1.</u> The 16 Negations Paradigm Conditional Inferences (Implicit Negation)[1].

	Referred Clause	Inferential Clause	Categorical Premise	Conclusion	Inference Type	Inferential Negation	Referential Negation	Double Negation	Matching
Explicit	A	2	A	2	MP1				*
Affirmation	A	NOT 2	A	NOT 2	MP2	*			*
	2	A	2	A	AC1				*
	2	NOT A	2	NOT A	AC3	*			*
Explicit	NOT A	2	A	NOT 2	DA3		*		*
Denial	NOT A	NOT 2	A	2	DA4	*	*	*	*
	NOT 2	A	2	NOT A	MT2		*		*
	NOT 2	NOT A	2	A	MT4	*	*	*	*
Implicit	NOT A	2	B	2	MP3		*		
Affirmation	NOT A	NOT 2	B	NOT 2	MP4	*	*		
	NOT 2	A	3	A	AC2		*		
	NOT 2	NOT A	3	NOT A	AC4	*	*		
Implicit	A	2	B	NOT 2	DA1				
Denial	A	NOT 2	B	2	DA2	*		*	
	2	A	3	NOT A	MT1				
	2	NOT A	3	A	MT3	*		*	

Only on the eight inferences that fall into the categories of *implicit affirmation* and *implicit denial* are the categorical premises represented by a negation. Therefore, if you manipulate explicit/implicit negation only these eight premises are affected. Those premises, which contain an affirmative categorical premise, remain unchanged irrespective of the explicit/implicit manipulation.

It was seen in Evans and Handley (1999) that implicit negation suppressed endorsements of the inferences. In fact, the matching bias effect diminishes substantially in the presence of explicit negation because explicit negation causes the production of more endorsements

[1] The column labelled 'inference type' contains the inference i.e. MP (modus ponens) but is appended by a number. This number represents which of the four negations-paradigm premises the inference is made up from. E.g. If the letter is an A then the number is a 2 (1), If the letter is an A then the number is not a 2 (2), If the letter is not an A then the number is a 2 (3), If the letter is not an A then the number is not a 2 (4).

of implicit affirmation and denial inferences. Referring to Table 3.1, if logic is controlled for, matching can be seen to be present if there are more endorsements to explicit affirmation inferences than implicit affirmation inferences, *and* more endorsement of explicit denial inferences than implicit denial inferences, thus creating an explicit/implicit referencing (categorical premise) divide. It will be seen from Table 3.1 that the categorical premise in explicit affirmation and denial problems, contains an item that is directly mentioned (matches) in the major premise, whereas the implicit inferences contain an item in the categorical premise that 'mismatches' the referred clause. This led Schroyens et al (2000) to discuss *mismatching* effects rather than matching effects as they saw the reduction in endorsements caused by implicit mismatching cases as the phenomenon of interest.

It is difficult to imagine that matching bias is due to processing difficulty with negations. When 'not' is added in the case of explicit negation, the matching effect disappears (i.e. more endorsements). Matching bias might thus seem to be a difficulty in *equating* an implicit negation with an explicit one i.e. being able to see a letter B as being representative of the clause 'not A'. Evans and Handley (1999) suggest, "matching bias...is due to suppression of the choice of the cases that implicitly negate the values in the conditional rule" (pg. 764) and that implicitly negated cases are just not seen as *relevant*. This is equivalent to Schroyens et al's (2000) label of mismatching bias. As there are only two options available for the materials in a conditional problem (matching or mismatching) it is difficult to decide just what the phenomena is: Do people generally accept inferences unless they are seen as irrelevant or do people generally deny inferences unless they are seen as relevant? The suggestion of *suppression* seems to indicate a general background strategy of endorsement. Either way, the findings still suggest a major role of relevance, be it heightened or suppressed.

Evans and Handley (1999) suggest that matching bias is directly linked to the linguistic forms of implicit /explicit negation. They offer a *double hurdle* theory to account for both implicit/explicit negation effects (matching bias) and double negation difficulties. The first hurdle is that the minor premise must be seen as relevant and this is facilitated if a negated minor premise is in the form of an explicit negation, otherwise they 'fall at this hurdle'. If the minor premise is still seen as relevant, then they may 'fall' at the double negation hurdle which involves an extra inferential step of falsifying a negative inferential clause.

Oaksford (2002) claims that negation effects are not due to linguistic heuristics as in Evans' (1998) heuristic-analytic theory of matching bias or even in Evans and Handley's (1999) more recent double hurdle theory of mismatching bias. Evans' account of negation difficulty is based on the assumption that the normal linguistic process of an explicit negation is to deny a presupposition rather than assert new information. That is, in the utterance, 'I did not watch television', the topic of the television still remains, rather than suggesting other activities that you could have been doing. When occurring in a conditional problem, the negative clause is seen as irrelevant unless it is portrayed explicitly. Oaksford on the other hand sees the role of negation as offering an alternative contrast set to the affirmative premise (outlined in Chapter 1). It is essentially the difference in size between the affirmative set and the negative set that causes difficulty with negation (the set defined by a negated clause is generally a lot larger than the affirmative one).

Oaksford (2002) claims that reasoners are more likely to endorse a conditional inference if the probability of the conclusion given the categorical premise is high, and this can explain differences found using explicit/implicit negation. Using an example of a DA inference, Oaksford claims that when given the premise, '*if A then 2*' the explicit version would consist of the minor premise and conclusion, '*not A*' therefore '*not 2*'. The probability of the conclusion given the minor premise in this instance would be higher than had the categorical premise been portrayed as an implicit negation ('*K*' therefore '*not 2*') e.g. P(*not 2*|*notA*)>P(*not 2*|*K*).

In Oaksford, Chater and Larkin (2000), probabilities were manipulated for both the antecedent and consequent in a conditional inference task (Exp.2). This was achieved by using a context in which cards were printed with a letter on one side and a number on the other. Probabilities (either high 'H' or low 'L') were manipulated by initial statements describing how many cards were printed with each letter and number. For instance, in their 'LL' (low/low) condition, participants were given information of the form; 'On the front of 10 of the 1000 cards there is an 'S'. and on the front of the remaining 990 cards there is a 'W'. Participants were then given the premise. 'If a card has an 'S' on the front then it has a '5' on the back'. This combination produces a low probability antecedent 'S' compared with the high probability 'not S' set (W). Four conditions were created; LL, LH, HL, HH. They found significantly higher endorsement rates for conclusions, which represented high probability sets. They conclude that the probability of the conclusion has a greater effect than the probability of the categorical premise.

For the selection task, Oaksford (2002) simply proposes that the use of explicit negations which all but eradicate matching bias, do so because of a matching strategy of the explicit negation that appears on the card (e.g. 'not a 2') to the clause represented in the major premise which might also be 'not a 2'. As pointed out in Chapter 1, this additional form of matching is referred to as *matching*$_2$ as opposed to the original matching phenomenon (now labeled as *matching*$_1$).

Set Size & Negation

Yama (2001) tried to disentangle the heuristic-analytic and set size accounts by manipulating set sizes along similar lines to Kirby (1994). Set size was manipulated in a selection task by giving participants the background knowledge that a computer prints cards with either a '+' or '-' on one side, and a number between 0-1000 on the other. One of four premises was then given; 'If a card has a +, then it (does not have) has a number from 1-1000' or 'If a card has a +, then it (does not have) has a zero. The cards on display remained the same for all four conditions; '+', '-', '0', '376'. If given a premise in which the consequent was 'a zero', the zero would represent a small set size. To falsify the 'zero' (FC) would require choosing a card that represented the large set size (376 represents the large set size of 1-1000). Likewise, to falsify the consequent 'a number from 1-1000' would require choosing the FC card which represented the small set size of 'zero'.

The findings showed a higher proportion of FC card choices when the FC set size was smaller, offering support for the optimal data selection account. The optimal data selection account proposes that reasoners see low probability set sizes as more informative. Thus a card depicting '0' has a lower probability of occurrence (1 in 1001) than a number which represents the set 1-1000 (1000 in 1001). The '0' is rarer and therefore more informative. Yama replicated these findings in his third experiment this time using lexically matching materials e.g. consequent clause being a number from 11-2000 with '2000' being depicted on one of the cards because in his first experiment there was a confound concerning matching and mismatching instances ('376' representing '1-1000'). In the third experiment, the matching bias effect still occurred together with a set-size effect.

Yama (2001) concludes that the phenomenon of matching is caused by *both* linguistically cued matching terms being seen as more relevant and optimal data selection accounts of set-size. Both accounts are seen in terms of relevance. Yama cites Sperber, Cara and Girotto's (1995) account of relevance theory in which they claim relevance is heightened if processing material causes greater cognitive *effects*. If on the other hand, more cognitive *effort* is needed to process the material, then its perceived relevance will diminish. The former is claimed to account for optimal data selection set-size effects whilst the latter is claimed to account for linguistically cued matching effects. Perceiving the probabilities offered by affirmative and negative clauses is seen to create differing cognitive effects, supposedly with smaller set sizes creating the largest cognitive effect because they offer more expected information in relation to the problem. Working out what is represented by an implicitly negated constituent on the other hand requires greater cognitive effort than that required by processing what is represented by the affirmative constituent. Therefore, the less cognitive effort involved in processing the affirmative clause makes it appear more relevant.

Relevance & Negation

MacDonald and Just (1989) found that explicit negation suppressed activation of the negated concept. They used sentences of the form, '*Almost every weekend, Elizabeth bakes no bread but only cookies for the children*'. Participants were required to read the sentence and then to respond to a probe word that followed on the screen. They had to respond by indicating whether the probe word had appeared in the previous sentence or not. So, for the above sentence, a following probe word could have been 'bread'.

They found then when an item had appeared negated in the sentence, response times were slower than had the word not appeared negated. They suggest the findings indicate that a negation suppresses the activation of the negated noun concept. Interestingly, initial encoding of the negation, measured by reading time, is not significantly different from non-negated concepts. This would indicate that comprehension of the negated concept is no more difficult than for an affirmative concept. Differences in response times using negations only become apparent when the task requires higher levels of processing.

Oaksford (2002) points out that the results of MacDonald and Just (1989) do not support Evans' (1998) contention that negations normally deny presuppositions and that the focus remains on the negated constituent. In MacDonald and Just's study, the presence of a negation shifted attention to the other named concept i.e. did not bake cookies, therefore baked bread. Although MacDonald and Just claim that negation shifts discourse focus, which in turn affects activation levels, in a conditional inference problem the major premise does not offer an immediate alternative to the negated concept. In MacDonald and Just's study, if Elizabeth bakes *no bread*, then the alternative would be that she would bake cookies. Only in the selection task is the alternative available as one of the card choices. If negation suppressed activation of the concept, then one would expect matching bias to disappear in the selection task. Reasoners would be inclined instead to choose the confirmatory response represented by the negated clause in the premise i.e. if given 'not 2' in the premise which would presumably suppress activation of the concept '2', the confirmatory response would be the choice of the '3' card.

A memory activation explanation of the selection task might be that, although activation of the mentioned concept is suppressed, it is nonetheless initially activated whereas its non-negated alternative is not. In the selection task, viewing the cards would then reactivate the trace activation initiated from reading the items in the conditional premise. If the activation of the concept were compounded (i.e. activated more than once with successive increments in the activation level), then activation levels for the negated concept would still be higher than the alternative card choice. For instance, given the premise, 'if A then not 2, the cards displayed might be; A, B, 2, 3. The A and 2 would be activated from reading the premise, although the activation of the '2' might be suppressed due to the negation. On viewing the cards, the 'A' and '2' will be activated again. This subsequent activation will be facilitated because of the initial activation that took place when reading the premise. This facilitation might lead to higher levels of activation of the 'A' and '2' cards above the activation level of the 'B' and '3'. This difference in amount of activation may lead the 'A' and '2' concepts to appear more relevant.

Lea and Mulligan (2002) took the role of negation further. Rather than studying the suppression of activation caused by an explicit negation, they wanted to find out what would happen if negations were *inferred*. For instance, given an opening statement, "according to the recipe, they could not use both fresh jalapenos and ground pepper in the chilli", (depicting the logical form 'not both A and B'), readers were then to discover that 'jalapenos' were to be used. The inference would follow that 'ground pepper' was not to be used and therefore represents an inferred negation. They found that reading times were slower when the inference that was given contradicted the logic of 'not both A and B' i.e. told that ground pepper was also to be used. They conclude that an inferred negation suppresses activation of that concept, thus making it less accessible.

Inferred negation occurs in conditional inferences when the process required is that of denial rather than affirmation. Referring to Table 3.1, in the explicit and implicit denial inferences, the categorical premise denies the referred clause. The inference then is that the inferential clause must also be denied by the conclusion. It could be that this explains the finding that denial inferences are endorsed less frequently overall, as their activation is suppressed by inferred negation. If these findings are taken into account with those of

MacDonald and Just (1989) who found that explicit negation suppressed activation, then this may explain the double negation effect. The double negation effect essentially involves the inferred negation of an explicitly negated concept i.e. not not A. From the results of MacDonald and Just (1989) and Lea and Mulligan (2002) it could be argued that there is a double suppression of activation effect. Before going into how these reasoning and relevance accounts of negation can be described in terms of memory representations, it will first be considered how negation affects memory retrieval. The following memory studies have looked at *both* negation and set size (in the form of categories).

Negation, Category Size & Memory

Sentence Verification Task

Early studies into negation difficulty used the experimental technique of sentence verification. Wason (1959, 1961) looked at the processing differences of affirmative and negative sentences, using four types of sentences of the forms:

2 is an even number (True Affirmative, TA)

2 is an odd number (False Affirmative, FA)

2 is *not* an even number (False Negative, FN)

2 is *not* an odd number (True Negative, TN)

Participants were asked to evaluate them with a response of true or false. There was also a separate *construction* version in which the number was omitted and participants were required to produce a number that satisfied the logic of the sentence. The overall pattern of response latencies of these studies and subsequent ones was of the order TA<FA<FN<TN with an assumption that a shorter response latency indicates a facilitation in mental processing. It will be noted that the two concepts mentioned in the sentences ('2' and odd/even numbers) could correspond to the relationship between the categorical premise and referred clause in the conditional inference negations paradigm (see Table 3.1). The

TA sentence verification task for instance, corresponds to the explicit affirmation inferences. For example, using an AC1 inference, 'If the letter is a vowel, then the number is even. The number is a 2. Therefore the letter is a vowel', a reasoner is required to deduce that '2 is an even number' in order to endorse (or produce) the conclusion. Along these lines, the following also concur: FN sentence & explicit denial inference; TN sentence & implicit affirmation inference; FA sentence & implicit denial inference.

What is of interest concerning the general pattern of findings with the verification task, and which has gone unnoticed in the literature up until now, is that there might also appear to be a *matching bias* effect as well as processing-negation difficulty. It can be seen that the two affirmative sentences (TA & FA) produce overall faster response latencies than the two negative types (TN & FN) and this is due to well established effects of negation already covered. But in addition to this, matching bias (in terms of a reduced response latency) can be seen when TA+FN>TN+FA which like the inference task, depicts a difference in response frequencies between explicit and implicit inferences.

Within the concept pairings it can be seen for instance that if the concept of '2' appears in the same sentence as the concept 'even number' (matching), it is responded to faster than if it appeared with the concept 'odd number' (mismatching) - whether the even/odd concepts are negated or not. If the concept of '2' appears in the same sentence as the concept 'not even', it is responded to faster than if it appeared with the concept 'not odd'. By comparing response times to TA/FA and FN/TN sentences, the presence of negations are controlled for, and the task requirement of either affirming or denying is counterbalanced. What is left is a pure matching effect, within an overall suppression of negation effect.

It will be remembered that in the selection task (i.e. 'if a vowel then an even number'), participants would be more inclined to select the '2' consequent card rather than the '3' card, irrespective of whether the concept 'even number' in the conditional statement was negated and also irrespective of the logic required. They would also be more inclined to endorse a conditional inference conclusion when the categorical premise and referred clause also reflected this matched pairing. Whilst several alternative reasons for the overall negation effect have been proposed such as the *emotionality* connected to negations, *extra steps* involved in comprehension and *linguistic factors* (See Evans, 1982, for a brief review), matching bias is also taking place.

Category Size & Memory

If negation causes the construction of contrast classes which are larger sets (and reflect higher probabilities), then studies looking at category size might reflect optimal data selection predictions. The memory studies that follow manipulated category size and measured processing difficulty in terms of reaction time. A similar type of structure to that of the sentence verification task was used in *categorisation reaction time* experiments as a method for studying memory structure. In categorisation reaction time experiments the focus was on how memory is organised semantically. They used the same format of sentences as in Wason's sentence verification task, but unfortunately did not generally use the full range of explicit negations. Mainly the studies looked at TA and FA sentences.

Landauer and Freedman (1968) used categories (e.g. dog) and category instances (e.g. collie) to test the general problem of how information is retrieved from memory. They found that response time was greater for larger categories than for smaller ones *only* when FA type processing was required (the words were presented separately with a task instruction requiring them to detect whether the target word was a member of the given category). They showed that participants were slower to falsify for instance, whether 'a tulip is an animal' than pairings in which the instance is not a member of a *smaller* category e.g. 'a tulip is a dog'. In their Experiment 2, using a different set of categories and category members, they showed again that larger categories produced longer reaction times to both affirming (TA) and denying (FA) responses. Positive (affirming) instances also produced overall faster reaction times than negative (denial) instances which could be accounted for by matching bias, although to make any substantive claims about this would require the systematic presence of negations within the materials.

Perhaps there is more represented by the materials than just their size, contrast class and the probabilities they represent. Collins and Quillian (1970) used the additional large category of 'living thing', but this time each instance was categorised twice i.e. Is a tulip a dog? Is a tulip an animal? With the belief that once people establish that a tulip is not an animal, they will be quicker to establish that it is not a dog (addressing their criticism of Landauer & Freedman, 1968). Again they found that 'dog' was easier to categorise than 'animal' for positive instances i.e. 'collie is a dog', quicker than 'collie is an animal' and also 'camel is an animal' quicker than 'camel is a living thing'. For negative instances (FA), they *partly* replicated the findings of Landauer and Freedman (1968) but only for the 'animal versus living thing' distinction. They conclude by stating, "it really does take longer to reject a negative instance as belonging to the larger category, but only sometimes".

The reason they give for the negative instance (FA) result only happening *sometimes* is given in terms of the semantic relatedness of categories and sub-categories. They suggest 'magnesium' is equally distinguishable from both categories of 'dog' or 'animal' and therefore elicits no difference in reaction times. But 'tulip' is quicker to identify as being 'not a dog' than 'not an animal' because somehow 'tulip' and 'animal' have a similar category relationship i.e. the superordinate category of 'living thing'. They therefore have a relationship in terms of the semantic structure of memory (semantic relatedness/semantically confusing). Such relatedness is postulated to stem from usage. Even opposite categories such as 'living' and 'non-living' can be semantically confused because of being "frequently contrasted". This makes it very difficult to judge any degree of semantic relatedness other than at an intuitive level. It would in itself, be a large study to produce any objective measure of this and even then would not encompass idiosyncrasies of knowledge attainment. The differences found in the reaction times on *positive* instances (TA) stems from having to make an extra inferential step to ascertain whether an instance is a member of the larger category i.e. collie is a dog – a dog is an animal. Therefore, the process takes longer to establish whether a collie is an animal, as opposed to whether it is a dog.

Although there is some further evidence of the semantic confusability hypothesis (Wilkins, 1971), Landauer and Meyer (1972) point out that this can not be the sole explanation of the difference in reaction times to small and large categories. Using data from Meyer and Ellis (1970) they point out that in this study, not only were negative instances used in the categorisation reaction time experiments, but also non-words (with English pronunciations). For instance, a non-word like 'mafer' was used as an instance concerning the category 'structures'. It still took participants longer to decide that 'mafer' was not a 'structure' than it did to decide that 'mafer' was not a 'building' (a smaller category than 'structure'). Hence, reaction times were affected by category size even though semantic relatedness between the categories and instances was dramatically reduced if not eradicated.

A further test that category size and not semantic relatedness is a major causal factor in obtaining reaction time differences, is cited by Landauer and Meyer (1972) from an unpublished replication of their 1968 study. Here, real words were used that differed in semantic distance from the categories. An objective measure of this was obtained by getting participants to rank the instances by "closeness of meaning" to the larger category. An example of this might be that 'tulip' is ranked closer to 'animal' than perhaps 'magnesium'. By dividing the data along the lines of 'closer words' and 'farther words' they found no significant difference between the two sets. Both sets were still equally affected by category size.

Category size was still believed to be a dominant factor in categorisation tasks up until studies into 'typicality' or 'relatedness' came to light (for a review see Chang, 1986). Smith, Shoben and Rips' (1974) study practically dismissed the idea that category size was a major contributing factor to reaction time differences by showing that it may have been an experimental confound. In their Experiment 1, they used two sets of *nested triples* i.e. collie/dog/animal. The two sets differed with respect to the 'relatedness' of the instance to either the smaller or the larger category. The relatedness measure was assumed from the production frequency norms obtained by Loftus and Scheff (1971). Here, ratings were obtained by asking participants to list superordinates for each of 50 instances. Thus, Smith et al used triples in Set 1 in which the *smaller* category was more frequently produced for the instance (therefore, they assumed a higher degree of relatedness). For Set 2 they used triples where the *larger* category was produced more frequently given the instance. They therefore manipulated relatedness between the two sets. They found that relatedness was more of a predictor in reaction time differences than category size, such that for the triple 'copperhead/snake/reptile' (Set 1) reaction times were smaller for identifying 'copperhead is a snake' than they were for identifying 'copperhead is a reptile'. In Set 2 the reverse was found. For instance in the triple 'chimpanzee/primate/animal', the 'instance/larger category' pairing was confirmed more quickly (chimpanzee is an animal).

McCloskey (1980) claimed that Smith, Shoben and Rips' (1974) study had actually confounded 'relatedness' measures with 'familiarity'. McCloskey, in a series of experiments, not only showed that Smith et al did not manipulate category size and instance-category relatedness independently, but when the effects of familiarity were controlled for, the Smith et al findings diminished. McCloskey showed that when participants had prior contact with the categories and hence were allowed to view them before the instance was displayed, they found (Exp. 2) that the relatedness effect was no longer significant. He found that comprehension times were longer on the Set 2, small categories used by Smith et al and this additional *comprehension* time explained Smith et al's findings. McCloskey points out that categorisation research using these 'natural' linguistic stimuli is problematic as to determining the true nature of semantic memory, as it is difficult to control for all extraneous variables across all conditions, but shows that 'familiarity' is a separate variable from measures of word frequency or word length. The question as to whether category size is still a predictor variable for reaction times remains open, especially concerning negative instances.

Although these tasks were mainly concerned with the processing of large and small categories, many of the studies did not specifically look at the presence or absence of a negation e.g. a tulip is *not* a dog and therefore, comparisons are difficult with respect to the study of negations in reasoning problems. What they possibly show is that there may be much more to the materials used in conditional problems, in particular, their representation and probability of occurrence using background knowledge and experience, rather than fixed 'encyclopaedic' knowledge. They suggest that the question of probabilities as described by the optimal data selection account may be more complex. Probabilities of the occurrence of an entity are likely to be stored implicitly in memory through experience, and be affected by such things as semantic relatedness, typicality and familiarity, just as these can affect category size effects. The above reaction time studies indicate that mental representations of category sizes are not fixed entities that corresponds to a thesaurus or encyclopaedia. It may be better to think of category size in terms of the amount of immediate mental associations one has with the concept mentioned. Oaksford (2002) concurs with this to an extent, by pointing out that a contrast class is a psychological and not a logical concept and that as well as context, can be 'computed' by a variety of sources of information such as phonetic, syntactic, semantic and pragmatic cues.

The concept of contrast class size may be more descriptive of the amount of mental association and subsequent activations connected to an item rather than just the probabilities they represent. It could be the sheer volume of connections that slows reaction time in the category size experiments and also causes difficulty in conditional reasoning tasks. It could be argued that the smaller spread of activation caused by smaller concepts increases their apparent relevance in reasoning tasks (rather than decrease their perceived probability), and facilitates processing as in the above memory tasks. Anderson's (1983) fan effect (Chapter 2, *Local Networks*) described how there is a finite amount of activation power that can be used for mental activation. If a concept activates many connected concepts and nodes then this spread of activation will reduce the activation peak for each connection and make processing the item slower. It was suggested that there may be a certain limit of activation, only above which, will concepts enter consciousness. Thus, if this finite amount of activation power is constrained to a smaller, semantic area (with fewer connections), then there is a greater chance that these nodes will have higher peaks of activation and will enter consciousness. If a concept 'pops' into consciousness, then it may account for it appearing more relevant. An activation account could explain both

Evans' (1998) matching account of heightened relevance and Oaksford and Chater's (1994) account of optimal data selection. Both of these accounts describe how matching bias occurs by a selection preference for the affirmative item. The affirmative item will have a smaller area of activation than the set described by its negation and might thus appear to be more relevant.

Memory Activation Model of Reasoning Biases

There is a developing parallelism between connectionist approaches and probability models of mental processes. Chater (1995) points out that there is a close relationship between neural network and inferential statistical models, saying, "The neural network architecture (defined purely by the pattern of connection between nodes) thus defines a family of probabilistic models, parametrized (sic) by the weights *w* associated with the connections" (Pg.212). Not only do Oaksford and Chater (1994) give a probabilistic rational analysis of the selection task and matching bias, but Anderson and Milson (1989) later developed Anderson's ACT* theory into a probabilistic rational analysis of human memory and then later (Anderson, 1991) human categorisation. If reasoning biases are caused by memory trace activations and general representational anomalies, then rather than distinguishing between current reasoning theories, any findings are more likely to unite them. With regard to the neural network/probability duality, whilst basic Baysian probability models of both memory and reasoning can be seen as fulfilling the need for simplicity (Chater, 1999) and support an account of human rationality, it still remains that reasoners get the problems wrong. Any notion of rationality accredited to biased reasoning processes may not stem from an adaptively rational attempt at solving the problem but from the adaptively rational mechanisms that represent the problem. It is hard to imagine that in the selection task, reasoners choose the '2' card because they see its rarity and rationally select it. It is more likely that reasoners aren't really paying much attention to what they may see as rather confusing problems. They are more likely to rely on a system of intuition – this intuition being the 'adaptively rational' aspect and one that seeks out 'relevant' aspects of a problem. The intuitive system here stems from the mental representation of the problem. It may be that the mental representation of a concept in terms of activation area is also synonymous with the probabilities represented by the concepts.

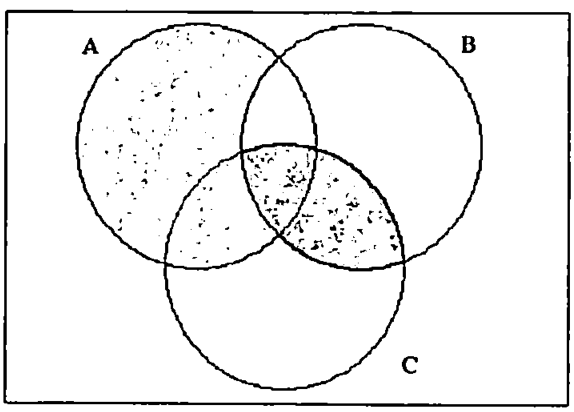
Various forms of mental representation have been offered concerning problem solving. One form of representation that could reflect a possible mental image of a problem is the Venn diagram. Johnson-Laird (1983) used Venn diagrams to represent syllogistic problems. A syllogism takes the form:

All A are B

No C are B

(therefore) No A are C

A Venn diagram is a schematic representation of the logical relationships of propositions. Johnson-Laird thus represents the above syllogism with a Venn diagram such as:



The shading indicates that the premises rule out these subsets. The lighter shading in the circle marked A shows that there are no circumstances in which an A exists without a B (all A are B). The darker shading in the area overlapping the B and C circles indicates the situation in which something cannot be both B and C (no C are B). The un-shaded portion of the remaining circles indicates the state of affairs that can now exist given the first two propositions. Thus, the conclusion 'no A are C' holds because there is no longer a representation of a situation in which A and C overlap (because the shading has ruled this out). Whilst the Venn diagram neatly illustrates an idea of how a mental representation may exist, Johnson-Laird points out that they are sophisticated mathematical notations and therefore have a limited application. In essence they are no more than the manipulation of truth tables. A further difficulty with representing the mind with such mathematical constructs is that they do not account for human psychology. It was pointed out previously that a negation is more a psychological construct than a logical one. If, for instance, the concept of 'cat' is negated, it is unlikely that this activates all mental concepts that aren't cats. It is more likely that such a negation activates concepts of a similar semantic domain e.g. 'animals'. A Venn diagram, or any other mathematical model, does not allow for such a restriction.

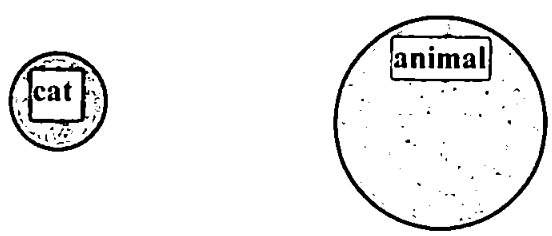
Whilst it is still useful to represent mental activation diagrammatically, a more psychological homespun version needs to be created in order to represent the complex interactions that may be taking place to represent a given problem mentally. It was earlier pointed out that recent memory theories postulate a distributed network of activations. These activations can occur with or without conscious awareness. It was suggested that the activation level of a concept has to reach a certain limit before it will enter consciousness. Also, there was the suggestion that activation power is of a limited capacity – a finite resource as it were, and that the more concepts that needed to be activated the more sparsely spread was the activation power. This spread also accounted for slower processing, as in Anderson's fan-effect. It was also noted that repeated activations changed the connections between concept and nodes, which thus facilitated further activation of that concept(s).

Any schematic mental representation of a reasoning problem must therefore take into account these psychological anomalies rather than just representing the problem in a mathematical and logical format.

The proposed mental activation model thus makes the following assumptions:

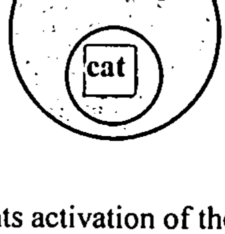
1. Smaller areas of activation have higher peaks of activation power and are quicker to process (fan effect).
2. A negation activates the named concept but at an inhibited level.
3. Repeated or overlapping activations increase activation levels (through changes in connections).

The overall assumption in the proposed model is that the levels of activation of a concept are directly related to the levels of perceived relevance. The first assumption (1) describes how a smaller area of activation may reach higher activation levels than a spreading activation involving larger concepts. This may happen because the same amount of activation energy is available for both large and small activations. Therefore, a small area of activation (e.g. 'cat') would be able to utilise the activation power available in order to increase the height of the activation of the concept(s). Higher levels of activation may enable the concept to enter consciousness and appear more relevant. The representation of large concepts (e.g. 'animal') would require the activation to spread out, thus not activating the concepts to as high a level. The activation of the two concepts of 'cat' and 'animal' could be modelled as:



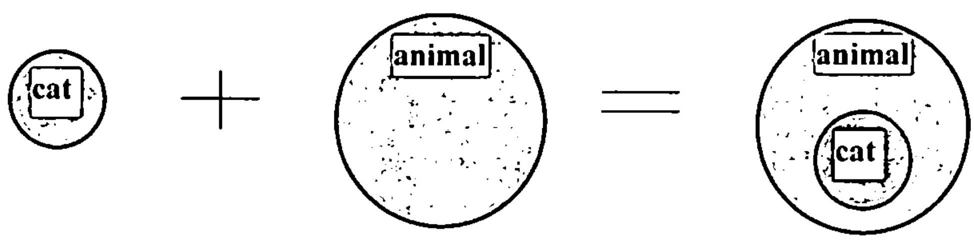
Where 'cat' evokes fewer concepts and so the area of activation is smaller, the total activation power remains constant and so the height of activation increases (indicated by darker shading). The concept of 'animal' evokes a larger spread of activation and thus the height of the activation peak is lower than that of 'cat'.

The second assumption (2) reflects the findings that a negation reduces or inhibits the activation of the negated concept (e.g. MacDonald & Just, 1989) and also that the presence of a negation slows reaction time on the sentence verification task. There are several ways of modelling this. The first one is that the concept alone is activated, but at a suppressed level. This suppressed level reduces perceptions of relevance. The second way of representing the mental activation of a negation is that the concept alone is not represented, but several similar concepts are activated instead e.g. 'not a cat' may be represented by the activation of several other concepts such as dog, mouse, budgie etc. This wider area of activation and subsequent lower peaks would explain reduced perceptions of relevance and create greater processing demands. Finally, a negation could be represented by an amalgam of the first two. That is, both the original concept and perhaps some of its contrast set concepts could be activated simultaneously, which again would spread the activation power available thus reducing overall perceptions of relevance. Two circles are therefore needed for the representation of just one concept 'not a cat':



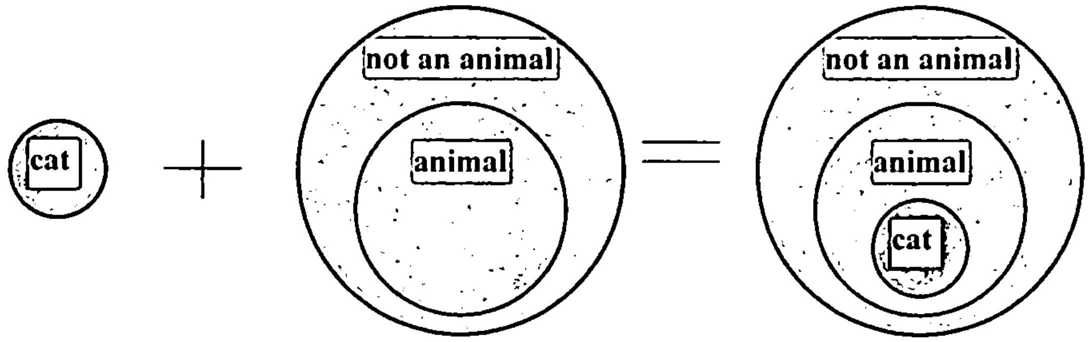
Where the inner circle represents activation of the concept 'cat' and the outer circle represents the extended activation of any further concepts which represent the contrast set. Very little is known whether the negation just suppresses activation or actually induces further activation of the contrast set concepts. Either way, representing the concept and negated concept(s) in two circles indicates that a larger area of activation is taking place due to the negation and thus lower levels of spreading activation reflect lower levels of perceived relevance. So, the concept that is negated is activated, but this activation power is spread to the activation of possible contrast class members. The named concept is therefore not seen quite as relevant than had it not been negated. Thus, the negated-concept mental representation also incorporates the notion of an inhibited level of activation.

The third assumption (3) that *repeated* activations increase activation levels can be represented by two overlapping circles that represent two activations of the same or similar concept. The final diagram represents activation of 'cat' and 'animal', which overlap in a semantic network.



Here, given the proposition that 'a cat is an animal', the concepts of cat and animal are activated and subsequently overlap in the semantic domain. The darker shading shows the increase in activation caused by the repeated activation of the same (or similar) concepts. Thus, given the sentence, 'a cat is an animal?' (TA) the two combined activations of animal and cat serve to increase the activation level of cat, therefore its perceived relevance is increased beyond that normally activated.

In the case where one of the two repeated concepts is negated such as, 'a cat is not an animal' (FN) the overlapping representation of 'cat' and 'not an animal' might be:



The spread of activation necessary to represent 'not an animal' may reduce perceptions of overall relevance and due to the fan-effect might explain differences in response times found on the sentence verification task, in that TA sentences are responded to more quickly than FN sentences. From the above diagrams, two mechanisms can be seen to affect the height of the activation peak: area of activation and activation overlaps. Both of these aspects are combined in the proposed model so that the overall area of activation is taken into consideration plus the amount of overlap between the concepts mentioned.

In the conditional inference task, it was seen that concepts are paired in the same way as the sentence verification task. These pairings occur between the categorical premise and referred clause, and there are also pairings from the conclusion and inferential clause. Looking at Table 3.2 (refer also to Table 3.1) for the 16 negations-paradigm conditional inferences it can be seen that there are only six possible concept pairing types (a-f):

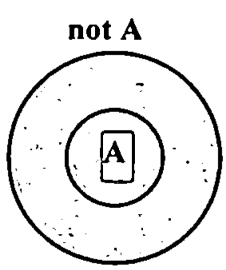
Table 3.2. Pairing Types for the 16 Negations-Paradigm Inferences.

Concept pairings	Categorical Premise/ Referred Clause	Conclusion/ Inferential Clause
a. A when A / 2 when 2	MP1,MP2,AC1,AC3	MP1,MP3,AC1,AC2
b. not A when not A / not 2 when not 2		AC3,AC4,MP2,MP4
c. A when not A / 2 when not 2	DA3,DA4,MT2,MT4	MT4,MT4,DA2,DA4
d. not A when A / not 2 when 2		MT1,MT2,DA1,DA3
e. B when not A / 3 when not 2	MP3,MP4,AC2,AC4	
f. B when A / 3 when 2	DA1,DA2,MT1,MT3	

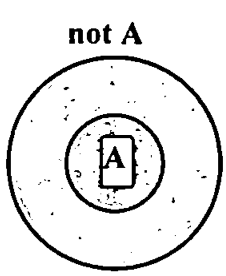
These pairings can be represented using five *pattern models* that reflect the individual activation sizes of the concepts and their corresponding activation overlaps (NB one model can account for both (c) and (d) as only their order is reversed). In constructing the five models only the antecedent representations will be used as examples, bearing in mind that 'A when A' will produce the same model as '2 when 2' etc. The five models are:

(1) 'A' when 'A'. Two small areas of activation that totally overlap.

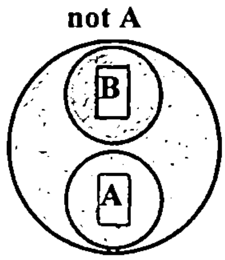
(2) 'not A' when 'not A'. Two large areas of activation that totally overlap.



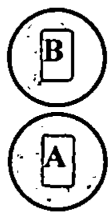
(3) 'A' when 'not A'. Small and large areas of activation that partially overlap.



(4) 'B' when 'not A'. A small and large area of activation that *may* partially overlap.



(5) 'B' when 'A'. Two small areas of activation that do not overlap.



Relevance here is assumed to stem from the height of the activation of the concept. It will be remembered that according to Evans (1983), relevance differences account for matching bias. This height of activation or peak can be affected by the total spread of activation (the amount of concepts activated) and the degree of overlap between the two concepts. To gain a more substantive indication of the differences in activation peaks between pattern models, some kind of numerical scoring of the models will be necessary. This can be achieved here by a crude method of scoring where; 3 is given for every small *area of activation*; 2 is given to every large area. Also, the proportion of *activation overlap* can be represented by; 3 for a complete overlap; 2 for a partial overlap; 1 for no overlap.

The points accumulated for the 'area of activation' caused by the two mentioned concepts, can thus be multiplied by the number given for the degree of 'activation overlap'. The higher the end figure, the higher the presumed peak of activation (and hence relevance) caused by the two concept pairings. The following figures can thus be calculated for the five models:

(1) 'A' when 'A' = (small area + small area) x complete overlap. (3+3)3 = **18**

(2) 'not A' when 'not A' = (large area + large area) x complete overlap. (2+2)3 = **12**

(3) 'A' when 'not A' = (small area + large area) x partial overlap. (3+2)2 = **10**

(4) 'B' when 'not A' = (small area + large area) x partial overlap. (3+2)2 = **10**

(5) 'B' when 'A' = (small area + small area) x no overlap. (3+3)1 = **6**

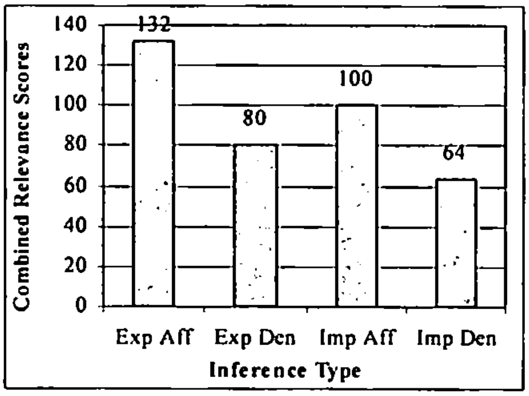
Each of the 16 conditional inferences creates two of these pattern models – one for the referred clause/categorical premise pairing and one for the inferential clause/conclusion pairing. To give an indication of the total relevance attributed to the inference, the scores for both models are combined to form a total *relevance score*. Table 3.3 shows the pattern model numbers and total relevance scores for each inference.

Table 3.3. Model Numbers and Associated Relevance Scores for each of the 16 Negations Paradigm Inferences using Implicit Negation.

Inference	Model Number		Relevance score		
Type	Antecedent	Consequent	Antecedent	Consequent	Total
MP1	1	1	18	18	36
MP2	1	2	18	12	30
MP3	4	1	10	18	28
MP4	4	2	10	12	22
AC1	1	1	18	18	36
AC2	1	4	18	10	28
AC3	2	1	12	18	30
AC4	2	4	12	10	22
DA1	5	3	6	10	16
DA2	5	3	6	10	16
DA3	3	3	10	10	20
DA4	3	3	10	10	20
MT1	3	5	10	6	16
MT2	3	3	10	10	20
MT3	3	5	10	6	16
MT4	3	3	10	10	20

If the 16 inferences are rearranged into the four categories of explicit/implicit affirmation/denial as in Table 3.1, and the relevance scores for each inference are combined, the following graph of total relevance scores can be obtained (Figure 3.1):

<u>Figure 3-1.</u> Accumulated Relevance Scores for Explicit/Implicit Affirmation/Denial Inferences (Baseline Model)



It can be seen from the total relevance scores in Figure 3.1 that the memory activation model not only encapsulates the matching bias phenomenon for conditional inferences but also the general pattern of responses observed in the literature. Matching bias is observed when there are more endorsements to explicit affirmation problems than implicit affirmation problems, and also more responses on explicit denial problems than on implicit denial problems. The presumption being that the more relevant an inference appears, the more likely it will be selected. The memory activation model also encapsulates the overall pattern of responses from conditional inference tasks using the 16 (implicit) negations paradigm inferences. Thus there may be seen to be a direct proportional relationship between the total relevance score and the generally observed frequency of endorsements given to the inferences.

The model at this stage is only a crude representation of possible relevance effects affecting endorsements given to conditional inferences at an unconscious System1 processing level (see Chapter 1, *Heuristic-Analytic & Dual Process*). It will be noticed that there are the same total relevance scores given to AC inferences as there are to MP inferences. Also there are the same total relevance scores given to MT inferences as to DA inferences. Previous studies generally show more MP than AC and more MT than DA. It must be noted that the model is only a *baseline* model of possible relevance effects at a perceptual System1 level, and does not take into account any attempt at actually thinking about the problem at a formal System2 level. The AC and DA inferences are actually incorrect, so if reasoners are trying to work out the problem (and a small percentage actually do get the problem right) then we might expect a certain proportion of fewer endorsements to the AC and DA inferences.

Double Negation

Whilst in its present form the model totally mirrors the matching bias effect, it does not account for the other response anomaly normally found in conditional inference responses – namely the double negation effect. Obviously, to get caught up in the double negation effect, reasoners must be trying to solve the problem (System2 processing) by withholding the given conclusion and trying to formulate their own, wherein they will encounter the second of the *double hurdles* (Evans & Handley, 1999). The present model assumes that reasoners are being totally led by the materials at a System1 level and endorse an inference only when the materials are *perceived* as more relevant. If reasoners are withholding the given conclusion to generate their own (System2), then this will change the activation overlap pattern model for the conclusion/inferential clause, whilst thinking harder about how the inferential clause fits in with the rest of the inference in order to generate their own conclusion.

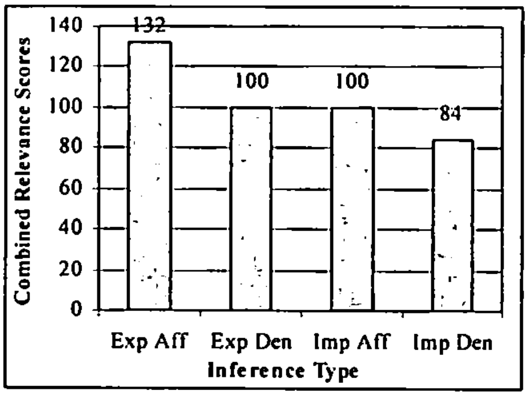
If this System2 thought process is incorporated into the overall model, then the representation of the inferential clause/conclusion activation needs to be slightly altered. Instead of the activation of the conclusion overlapping with the inferential clause it is removed completely. The assumption here is that reasoners will think harder about the inferential clause only. This extra thought could be represented by an additional activation of the inferential clause concept. So according to the mechanisms for working out the relevance score, it could be claimed that the inferential clause is totally overlapped by itself and therefore its relevance increased. These new models and figures (no activation of the conclusion/repeated activation of the inferential clause) are given in Table 3.4. It can be seen for example, that the original model for a DA1 conclusion/inferential clause activation was model number 3. This represented the activation of 'A and not A' (or '2 and not 2'). If the inferential clause alone is reactivated without the conclusion (e.g. '2 and 2'), it is now represented by model number 1, and the relevance score increases from 10 to 18.

Table 3.4. Revised Model Numbers and Associated Relevance Sores for each Inference.

Inference Type	Model Number Antecedent	Consequent	Relevance score Antecedent	Consequent	Total
MP1	1	**1**	18	**18**	36
MP2	1	**2**	18	**12**	30
MP3	4	**1**	10	**18**	28
MP4	4	**2**	10	**12**	22
AC1	**1**	1	**18**	18	36
AC2	**1**	4	**18**	10	28
AC3	**2**	1	**12**	18	30
AC4	**2**	4	**12**	10	22
DA1	5	**1(3)**	6	**18**	<u>24</u>
DA2	5	**2(3)**	6	**12**	<u>18</u>
DA3	3	**1(3)**	10	**18**	<u>28</u>
DA4	3	**2(3)**	10	**12**	<u>22</u>
MT1	**1(3)**	5	**18**	6	<u>24</u>
MT2	**1(3)**	3	**18**	10	<u>28</u>
MT3	**2(3)**	5	**12**	6	<u>18</u>
MT4	**2(3)**	3	**12**	10	<u>22</u>

The bold print indicates the model type and relevance score for the inferential clause, which is now *reactivated* rather than the conclusion being activated. The concept that forms the conclusion is no longer activated in this model. The parenthesis indicates the previous model number (Table 3.3) in which both the inferential clause and conclusion were activated. It can be seen that this new amendment to the model does not change the relevance scores for MP and AC inferences. Figure 3.2 shows how this amended model is incorporated into explicit/implicit affirmation/denial inferences.

Figure 3-2. Accumulated Relevance Scores Without Conclusion Activation for Explicit/Implicit Affirmation/Denial Inferences (Reactivation Model).



The model still shows matching bias, in that the explicit affirmation relevance score is higher than the implicit affirmation score and the explicit denial score is higher than the implicit denial score. The only differences in the pattern of scores involve the DA and MT inferences. The double negation effect is apparent when (DA1+ MT1)>(DA2+ MT3), and (DA3+ MT2)>(DA4+ MT4). In Table 3.3 (baseline model) all these pairings had the same scores – they were equivalent rather than showing any differences; 32 *(DA1+ MT1)*=32 *(DA2+ MT3)*, and 40 *(DA3+ MT2)*=40 *(DA4+ MT4)*. In Table 3.4 with the amended relevance scores it can now be seen that the activation model reflects a double negation effect 48 *(DA1+ MT1)*>36 *(DA2+ MT3)*, and 56 *(DA3+ MT2)*>44 *(DA4+ MT4)*. So if the activation of the conclusion concept is replaced with the *reactivation* of the inferential clause concept, the memory activation model now fits exactly with the findings in the literature with regard to the conditional inference task.

The concept of memory activation explains both relevance effects with regard to matching bias and now seems to imply relevance (activation height) effects with regard to the double negation effect. Both models are still useful with regard to reasoning in that it could be argued that not all reasoners really attempt to formulate their own conclusion. Resembling Evans and Handley's (1999) double hurdle theory, it could be argued that if reasoners do not attempt to solve the problem, then they are led solely by the activation of *all* the materials used in the task – *baseline model* (first hurdle). If reasoners do try to attempt to solve the problem, then they do not represent all the materials, deciding instead to suppress the given conclusion in order to generate their own (*reactivation model*). This attempt at reasoning still succumbs to activation biases and they fall at this second hurdle. Interestingly, the model predicts the double negation effect, not through difficulties with the denial of an explicit negative. It can be seen that the double negation effect could be occurring just by the representation of the materials. What is being processed in the model is not the inferred denial of an explicit negative, but just its reactivation in terms of memory processes.

Explicit & Implicit Negation

The use of explicit negation diminishes the matching effect to the point that there is no longer found to be significant differences in the proportion of responses given to matching than mismatching categorical premise/referred clause pairings. Matching pairings are seen to be those of the form 'A when A' and 'A when not A' whereas mismatching pairings are those of the form 'B when A' and 'B when not A'. As was pointed out earlier, the use of implicit negation only affects implicit affirmation and implicit denial inferences. If explicit referencing is used in the activation model, the activation patterns for the categorical premise/referred clause pairings have to be changed. For instance in the MP3 inference the pairing was previously represented by pattern model number 4 (B when not A). This now has to be changed for explicit negations to pattern model number 2 (not A when not A). This increases the relevance score from 10 to 12. The other implicit affirmation inferences change their scores the same way. (MP3, MP4, AC2, AC4 all change from 10 to 12). Likewise, all the previously *implicit* denial inferences change their pattern models for the categorical premise/referred clause from pattern model number 5 to 3, with their scores changing respectively from 6 to 10 (DA1, DA2, MT1, MT3).

If these new relevance scores are placed in the *reactivation model*, the resulting relevance totals are; Explicit affirmation 132, Explicit denial 100, Implicit affirmation 108, Implicit denial 100. The explicit affirmation and explicit denial scores remain unchanged (see Table 3.4). The largest difference has happened to the implicit denial inferences whose total has now gone up from 84 to 100. The implicit affirmation scores have gone up from 100 to 108. These new scores have shortened the gap between explicit affirmation/denial and implicit affirmation/denial which is normally an indicator of matching bias. The revised explicit negation model now does not predict any matching bias difference between explicit and implicit denial inferences (both now achieve scores of 100). There still remains a predicted difference between explicit and implicit affirmation type inferences (132/108). Overall, the matching bias effect is predicted to decrease substantially.

The Selection Task

The only effects to be modelled in the selection task are matching bias and possible contrast size effects (there are no double negations as in the inference task). It is widely observed that the biggest matching bias effect occurs with the consequent card choices because reasoners generally accept that the topic of the conditional is reliant on the presence of the antecedent clause. This so-called 'if-heuristic' cues detection of the antecedent condition, otherwise the conditional is seen as irrelevant. The standard abstract selection task in which the antecedent and consequent items are mirrored on two of the four cards will be used as an example. A standard task of the form 'if the letter on one side is an A then the number on the other side is a 2', has the cards on display depicting; A, B, 2, 3. If the negations paradigm is applied it creates four conditionals of the form; If A then 2; If A then not 2; If not A then 2; If not A then not 2.

Using only the consequent numbers for examples (the antecedent models have the same patterns), the activation of the concepts mentioned on the cards and in the premise are; 2 when 2; 2 when not 2; 3 when 2; 3 when not 2, thus creating four possible activation pattern models. Again, using the relevance scoring method applied for conditional inferences, the following scores are calculated; [2 when 2] (3+3)3=**18**; [2 when not 2] (3+2)2=**10**; [3 when 2] (3+3)1=**6**; [3 when not 2] (3+2)2=**10**. Matching bias is seen when there are more true consequent (TC) '2 when 2' cards selected than '3 when not 2' cards are selected (**18>10**) and more false consequent (FC) '2 when not 2' cards than '3 when 2' cards (**10>6**). The activation model does not represent the '3' card as a larger set as in the optimal data selection account but even if it did, it does not drastically alter the model or its matching predictions, changing the score for '3 when 2' from '**6**' to '**10**', and '3 when not 2' from '**10**' to '**12**'. Thus, according to the activation model, it does not matter whether reasoners re-code a '3' as representing the contrast set of 'not a 2', and therefore supports a relevance rather than a contrast class account.

To test the model further requires manipulating the activation area of the concepts mentioned in the selection task. Yama (2001) did this in his selection task materials investigating the optimal data selection account of matching bias. As was pointed out earlier, Yama created a scenario in which a computer printed cards with a '+' or '-' on one side and a number from 1-2000 on the other side. Only the consequent was manipulated providing four negations paradigm clauses of 'a number from 11-2000', 'a number from 1-10', 'not a number from 1-10', 'not a number from 11-2000', with the consequent cards depicting '10' and '2000'.

The activation model can represent these concepts as '11-2000' as a large area of activation and '1-10' as a small area of activation. Rather than make encoding of the relevance scores too complicated, only the concepts of large and small areas of activation will be considered together with partial or no activation overlap. For the model, only the *relative* activation area sizes are important. All overlaps using categories only *partially overlap* in activation (e.g. 'dog' only partially overlaps the concept of 'animal'). If it is considered that the following concepts concerning the small '1-10' set are all small areas of activation and all the concepts concerning the large set '11-2000' are large areas of activation then the following concepts can be encoded; '10', '1-10', 'not 1-10' all score 2 (small area of activation); '2000', '1-2000', 'not 1-2000' all score 1 (large area of activation). The concepts shown on the cards of '10' and '2000' are in the normal world, of equal activation size. In Yama's task they are cued by the scenario to represent only one of two sets – small or large and therefore Yama's experimental manipulation is deemed to have been successful with regard to what sets the cards represent. So, in this activation model they are encoded as small and large areas of activation accordingly. Also, whilst the presence of a negation increases the activation area according to the model, the concept of 'not 1-10' would still be seen as activating a smaller area than the concept of 'not 11-2000'. The model does not account for re-coding the explicitly negated premise set into the alternative affirmative set e.g. 'not 11-200' is not re-coded into '1-10'.

Thus, four models can represent the problems:

(1). If + then 11-2000	TC '2000' = (large+large)partial overlap = (1+1)2 = **4**
	FC '10' = (large+small)no overlap = (1+2)1 = **3**
(2). If + then 1-10	TC '10' = (small+small)partial = (2+2)2 = **8**
	FC '2000' = (small+large)no overlap = (2+1)1 = **3**
(3). If + then not 1-10	TC '2000' = (small+large)partial overlap = (2+1)2 = **6**
	FC '10' = (small+small)partial overlap = (2+2)2 = **8**
(4). If + then not 11-2000	TC '10' = (large+small)partial overlap = (1+2)2 = **6**
	FC '2000' = (large+large)partial overlap = (1+1)2 = **4**

From the model it can be seen that matching bias is still predicted, in that the accumulated relevance scores for the matching cards are greater than those for mismatching cards ((4+8+8+4)>(3+3+6+6)). Also, the model explains why there appears to be a preference for selecting the small FC set cards over the larger ones. The *small* FC scores combined produce a higher relevance score than the *large* FC scores (3+8)>(3+4).

Predictions

It has been shown that a model of conditional reasoning based on mental activation can assimilate the major findings that have stemmed from the majority of studies on the conditional inference task and selection task: matching bias; double negation; explicit/implicit negation. The model now has to make unusual predictions concerning conditional problem responses. If the response patterns found on standard conditional inference tasks are caused by memory activation anomalies, then one might suspect that manipulations that have been found to affect memory tasks might also affect reasoning tasks. It was seen that early memory studies used category sizes in a sentence verification type task. It was found in these studies that increasing the category size of a concept also increased its response time. So that, if given the proposition 'a cat is an animal', this may be quicker to verify than if given the proposition 'a cat is a living thing'. The concept of 'animal' in this instance being smaller than that of 'living thing'.

Looking at both the baseline (System1) and reactivation (System2) models, they predict that the larger the spread of activation, the lower will be the activation peak. Using the same materials throughout all 16 problems, the models do not reflect any other response patterns than those already accounted for. But, if two or more sets of the negations-paradigm inferences were created which differed in the amount of activation area they produced, the model would predict differences in response patterns. Large categories for instance would create larger areas of activation and therefore lower activation peaks than smaller categories. As in the category reaction time memory studies, the activation models would predict that the larger the spread of activation caused by processing the concept, the less relevant it would be perceived and more time would be needed to process or mentally represent it (fan effect). Therefore, the less relevant it is perceived, the less likely a conditional inference with these materials will be endorsed.

The following experimental chapters thus outline studies in which categories were used in the selection task to see if matching bias still occurs and also if all the card concepts can interfere with memory activation. Category size was manipulated in the conditional inference task to further test the mental activation model. The general prediction being that the larger the area of activation necessary to represent the concepts in a problem, the less relevant they will seem and therefore they will elicit fewer endorsements. This will particularly result in a decline in the amount of observed matching bias. *Nested triples* of categories and category instances will be used. A nested triple takes the form for instance, 'cat/animal/living thing. They are nested in that a cat is a smaller concept than animal, which is again smaller than living thing. They are all semantically related, which guarantees that their relative sizes are in line with the intended manipulation. For instance, it is safe to assume that cat is a smaller category than animal because it is a subset of animal. Whereas it could not be guaranteed that cat is a smaller concept than bird (there may be just as many types of cat represented in someone's memory, as there are birds).

It will be shown in the experimental chapters that the two models (baseline & reactivation) predict more matching responses (endorsements) when smaller concepts are used in the problems than larger concepts. So that the parings of 'cat/animal' will elicit more responses than the pairings of 'cat/living thing', with the 'pairing animal/living thing' eliciting the fewest endorsements. No other theory of reasoning at this time predicts a response pattern of this nature using these categorical materials.

Chapter 4

Experiments 1, 2 & 3 – Selection Task

The following studies are an investigation into the general pattern of 'matching bias' that has been well documented concerning a popular reasoning paradigm – the Wason selection task. The aim will be to clarify what aspects of the materials used in the task actually produce the matching bias phenomenon. Matching bias on the selection task is seen to be a response preference for items on the cards that are named in the premise, irrespective of negation or logical requirements of the task. For instance, a standard *implicit* abstract selection task may contain the premise, 'if there is an A on one side there is a 2 on the other', with cards displaying, 'A', 'B', '2', and '3'. Reasoners typically choose the 'A' and '2' cards, whereas the logically correct falsifying choices would be the 'A' and '3' cards. The card items that mirror the items in the premise are seen as more relevant in ascertaining the truth or falsity of the premise. With explicit negation, all the cards on display represent the topic items in the premise for both the antecedent and consequent clauses e.g. using the above example, the 'B' card would be replaced with the words 'not A' and the '3' card with 'not 2'. When the cards are represented in this way, the matching bias effect disappears. It would seem therefore, that when there is a direct lexical link between the cards and the premise, the cards are seen as more relevant and are more likely to be selected. The converse here being that when items on the cards do not lexically match the items in the premise, they are seen as irrelevant to the task.

The question arises as to what degree of relatedness between the items in the premise and the items on the cards, has to be in existence for the cards to be perceived as relevant to the task? Matching has never been fully defined – do items match on a lexical, word-for-word basis or do they match because one item represents the other and is therefore logically equivalent? The findings using explicit negation seem to indicate that the materials have to be lexically matching in order to be seen as relevant. When an implicit negation is used on one of the cards, it no longer matches the topic named in the premise. Using the above premise, the implicitly negated consequent card choice is a '3' whereas the explicitly negated card would display 'not a 2'. The 'not a 2' card is more likely to be selected (Evans, Clibbens & Rood, 1996). Reasoners therefore would not seem to see the relevance of the implicit negation possibly because it does not lexically mention the original referent. The '3' and 'not a 2' cards still have the same *referential* meaning in the context of the problem though and still fulfil a logical function of the task, but the implicitly negated card no longer appears to be relevant. Could matching bias just be a direct result of lexically matching terms?

Wason Selection Task

There is a long tradition of reasoning studies involving the Wason selection task dating back from the late 1960's. Over the years, various manipulations have been made with the content of the materials used in the task. Typically, these have fallen into two main content type categories: 'abstract' and 'thematic' (outlined in Chapter 1, *Realistic/Thematic Material*). The use of thematic or real world materials in the selection task usually eradicates the matching bias effect possibly because the materials used heighten the perceived relevance of all the cards through pragmatic cues or other background knowledge. Differences in matching effects between thematic and abstract tasks can not be entirely due to differences in task presentation with regard to lexical matching although it is important to understand what the underlying mechanisms are that cause matching bias in abstract tasks. Abstract tasks offer an opportunity to investigate reasoning biases in a more controlled way, reducing the influence of pragmatic factors such as prior belief. Such an understanding may offer a better insight into the complexities that arise in thematic or everyday reasoning tasks.

The original Wason task (1966) has been one of the only studies that have not used lexically matching cards. The original task used letters and numbers to form an abstract deductive reasoning task. Reasoners were told that the cards had letters printed on one side and numbers on the other. Given the premise, 'if there is a vowel on one side, there is an even number on the other', the cards displayed were, 'A', 'K', '2', '7'. Wason found that there was a general response bias to select cards that represented the items mentioned in the premise. Reasoners were more inclined to select the 'A' and '2' cards which Wason believed to be a form of verification or confirmation bias.

Negations Paradigm

The systematic manipulation of negations throughout a conditional premise has been coined 'the negations paradigm' (Oaksford & Stenning, 1992). If negations are introduced into the Wason selection task, it effectively creates four possible conditional statements, the subsequent logical responses to which are also respectively changed. For instance, if given the conditional premise, 'if there is an A on one side then there will be a 2 on the other side', according to propositional logic, the responses that may refute this rule are the selection of the 'A' and '3' cards. If the conditional premise contained a negation such as 'if A then not 2', the correct, logically refuting choices would now be 'A' and '2'. The four possible premises that can be constructed are:

If A then 2

If A then not 2

If not A then 2

If not A then not 2

Matching Bias

There is found to be a general matching bias response pattern across all four of these negation paradigm premises, irrespective of the different logical requirements created. Evans (1998, for a review) postulated that a response bias was formed due to the materials themselves. Matching bias seems to be generally unaffected by the logical structure of the problem and also the negations used. In seeing any of the four premises outlined above, reasoners have a tendency to still choose the 'A' and '2' cards irrespective of negation and subsequent logical requirements of the task. The general indicator of matching bias can be seen in the selection task by controlling for logic when there are more '2' card selections when the premise states 'there is a 2', than '3' card selections when the premise states 'there is not a 2'. Also, matching bias is indicated when there are more '2' card selections when the premise states 'there is not a 2', than '3' card selections when the premise states 'there is a 2'. Comparisons between the materials along these dimensions control for the logical requirement of the task.

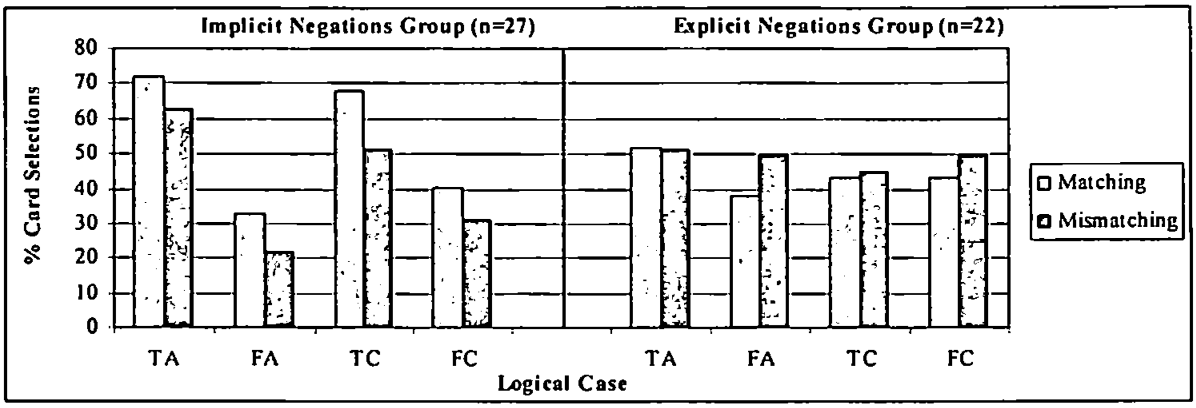
The general pattern of responses to the abstract selection task using implicit and explicit negation are typified in Figure 4.1. The TA, FA, TC and FC abbreviations describe whether logically the card represented the True/False Antecedent/Consequent clause. For instance, given the antecedent clause, 'if there is not an A...' the TA card might be 'B' (not an A) and FA card might be 'A'. Examples of the implicit and explicit negations used in that study are as follows:

If the letter is R then the number is not 3

the letter T (Implicitly negated antecedent card)

a letter which is not R (Explicitly negated antecedent card)

Figure 4-1. Typical Abstract Selection Task Findings Using Implicit and Explicit Negation. (Taken from Evans, Clibbens & Rood, 1996, Exp.2).



Using selection frequency data from Evans, Clibbens and Rood (1996) it can be seen that there is an overall response preference for matching rather than mismatching cards in the *implicit* group irrespective of logical case. There also tends to be a trend of confirmation bias in that there are more TA and TC card selections overall. There is no evidence of matching bias in the *explicit* group.

Overview of Current Matching Bias Accounts

Heuristic-Analytic Theory

Evans (1998) claims that the underlying mechanism to matching bias may be caused by the materials in the premise being seen as more relevant to the task than alternative choices and somehow take on a heightened importance when selection decisions are being made. The heuristic-analytic theory outlines how preconscious heuristics affect selections made in the task. There are two aspects to the initial heuristic stage of reasoning. These are the 'if' and 'matching' heuristics. The 'if' heuristic describes how the linguistic nature of the word 'if' directs attention to true antecedent cases (the linguistic topic of the premise). Reasoners are more likely to see the detection of this clause as relevant to the premise. The 'matching' heuristic describes how reasoners' attention is drawn to the affirmative value of the given concept irrespective of negation. The reason for this is given in terms of the linguistic function of negations. The presence of a negation in everyday use still directs attention to the negated constituent rather than the set of circumstances that the negated clause represents. Therefore, the prediction is that more affirmative cases will be selected irrespective of negation or logical case.

Mental Model Theory

Johnson-Laird and Byrne's (1991; 2002) mental model theory outlines how matching bias may occur because of initial representational anomalies of the mental model. Matching bias occurs because the initial model explicitly represents the affirmative concept but it may or may not be *flagged* by a negation marker depending on the polarity of the premise clause. Demands on working memory caused by the amount of models that can be constructed from a premise cause processing difficulties which can result in the affirmative items being perceived as being relevant to the task, whilst implicitly negated 'mismatching' items are seen as irrelevant. Evans (1993) has pointed out that this unconscious process of selecting relevant aspects of the task coincides with the heuristic stage of the HA theory.

Optimal Data Selection

The tendency to choose cards that represent the affirmative form of the concept mentioned in the premise has been classed by Oaksford and Chater (1994) as a rational attempt at optimising the data that can be obtained from the task. Here, the affirmative card choices represent smaller probability sets than the cards that represent the negated or contrast set. For instance, being told that cards have letters on one side and numbers on the other, and given the premise, 'if there is an A on one side there is a 2 on the other', the '2' card represents a probability of occurrence (in a set of single numbers) of 1 in 10. The card which represents 'not a 2' implicitly e.g. '7', represents a contrast set in which the probability of occurrence is 9 in 10. The smaller probability set ('2') is more informative in that it represents a rarer occurrence, and therefore it is more likely to be chosen in order to gain information.

Memory Effects as an Alternative Matching Bias Account

A more specific account of the underlying mechanisms that cause matching involves spreading or distributed memory activation theories. The mental activation model put forward in Chapter 3 shares similar theoretical assumptions that have been offered as explanations for memory priming phenomenon. One theory behind priming effects is that put forward by McClelland and Rumelhart (1986). They proposed that exposure to a stimulus or concept produces a pattern of connected neuronal activations. Each activation carries with it a form of 'incremental' learning in that every time connections between nodes in the distributed network are made, an alteration occurs in the connections that accounts for a type of learning process. Subsequent activations of the same mental pattern thus become facilitated.

Repetition priming effects (Chapter 2) can be very durable, lasting in some cases for several days between prime and target exposure. Semantic priming effects (i.e. using terms that are not lexically identical, but share the same meaning) on the other hand have been generally found to be short lived, surviving only long enough to record an effect in the immediate experimental task. If priming effects occur in the selection task and are responsible for matching bias then it is not apparent as to what type of priming is occurring. One indicator may be that the use of explicit negation increases choices of cards that represent the negation (Evans, Clibbens & Rood, 1996). It could be argued that explicitly negated cards also contain the lexical item that appears in the premise. For example, given the premise, 'if there is an A on one side, there is not a 2 on the other side', an explicitly negated consequent card would depict 'not a 2'. This lexical match would appear to increase the card's perceived relevance. With an implicitly negated card (e.g. '3') there is no longer a lexical match with the premise clause. The '3' card only matches in terms of reference in that '3' is 'not a 2'. It would seem that if priming is occurring in the task, and is responsible for certain cards appearing more relevant, then it may be likely that repetition priming, which involves the direct lexical match between prime and target, is responsible.

Lexical & Semantic Matching

Until now, much of the selection task research into these biasing effects has used the same lexical material in the premise as there appears on the cards. There is much in the way of memory literature on the long-term carry over effects of repetition priming and it may be that some memory effect of this nature might offer insight into the notion of heightened relevance when making card selections. One way of manipulating this lexical link between the premise and the cards whilst keeping the facets of the problem logic intact, is to use categories. Problems could be created where the premise items do not lexically match the card items. This can be achieved by using categories and category inclusions. Two forms of the selection task can then be created; one in which the items mentioned for the antecedent and consequent clause in the premise lexically match the items shown on the cards e.g. 'cat' and 'cat'; and one in which the items in the premise are superordinate categories of the items mentioned on the cards e.g. 'animal' and 'cat'. If repetition priming was responsible for matching effects, then by making a task in which the items no longer lexically match may show a reduction in matching responses.

Card Probabilities Vs Mental Activation of Concepts

The mental activation model clearly specifies the underlying mechanisms that may be responsible for the matching bias effect. The only current theory that offers such specificity is the optimal data selection account (ODS) of Oaksford and Chater (1994). On the surface both theories make similar predictions, but a closer inspection reveals some minor, though notable differences. The ODS account predicts differences in selection frequencies between the lexical and semantic tasks. The ODS account claims that cards represent probabilities with specific reference to the clause mentioned in the premise. Such that given the premise, 'if there is an A on one side there is a 2 on the other side', and given the consequent cards '2' and '3', reasoners will see the '3' card as representing the set 'not a 2'. According to the ODS account, the probability of the '3' in the world of numbers is not taken into account. Otherwise the probability of a '3' would equal the probability of a '2'. Instead, the '2' card is seen as representative of the premise clause ('2'), whereas the '3' is seen as representative of the contrast set ('not a 2'). Therefore, the probabilities represented by the two cards are now disproportionate in that the probability of a number 'not being a 2' is higher than the probability of it being a '2'.

Yama (2001, - see Chapter 3, *The Selection Task*) manipulated the probabilities of the consequent card selections between tasks and found that participants were more inclined to choose an FC card when it represented a low probability set. The original Oaksford and Chater (1994) ODS account described how these probability differences *within* a selection task explain the matching effect of a bias towards choosing the affirmative matching card because it represented a smaller probability of occurrence than the implicitly negated alternative. The affirmative consequent card was therefore seen as more informative as it represents rare information and this explained matching bias effects within the selection task. Oaksford (2002) also uses Yama's findings of differences in FC card choices *between* tasks as support for ODS account.

In the present study only the items mentioned in the premise are changed between conditions – the card items remain the same. For the premise in the *semantic* condition, a superordinate category of the item mentioned on the cards is used e.g. the premise item might be 'tree' and the matching card would display 'oak'. In the lexical condition, the premise item is lexically identical to the matching card ('oak'). Two types of premise can thus be formed e.g. 'If a card does not have a fish (cod) on the left side, then it has a tree (oak) on the right side', with the items in parentheses indicating the terms used in the lexical condition. In both cases the cards on display would show, 'cod', 'fox', 'oak' and 'ivy'. Using category superordinates in the semantic condition in this study should change the probabilities *represented* by the cards. Even though in the present study the cards on display remain the same between tasks, the use of superordinate categories in the premise changes what the cards represent. For example, in both the lexical and semantic conditions the consequent cards might depict 'oak' and 'ivy'. In the lexical condition the consequent clause could be either '...then it has (does not have) an *oak* on the right side' or in the semantic condition '...then it has (does not have) a *tree* on the right side'.

According to the ODS account, the cards in the lexical condition would represent 'an oak' and 'not an oak', and in the semantic condition they would now represent 'a tree' and 'not a tree'. The probabilities represented by the cards are now different between tasks. The probability of something being an 'oak' is lower than the probability of something being a 'tree'. Conversely, the probability of something 'not being an oak' is now higher than the probability of something 'not being a tree'. Within the lexical task, the *difference* in probabilities between the two consequent cards is now greater than the difference in probabilities between the two consequent cards in the semantic condition. Thus, according to the ODS account, there are likely to be more choices of an FC card (using affirmative premises), in the semantic condition than in the lexical condition (there is an interaction of negation as will be shown in Table 4.1).

According to the predictions of Oaksford, Chater and Grainger (1999, Pg. 200), "there should be effects of P(*q*) (i) on the *q* card such that there are more *q* card selections when P(*q*) is low, and (ii) on the *not-q* card such that there should be more *not-q* card selections when P(*q*) is high". These predictions can be broken down by logical case (TC & FC selections). For the lexical and semantic conditions in this study, these predictions for the negations paradigm are best outlined in Table 4.1.

Table 4.1. Optimal Data Selection Predictions for Low Probability Q (lexical condition) and High Probability Q (semantic condition) Card Selections.

Premise Clauses	Lexical – *TC* *(Low P[q])*	Semantic – *TC* *(High P[q])*	Lexical – *FC* *(Low P[q])*	Semantic – *FC* *(High P[q])*
If P then Q	More 'Q'	Less 'Q'	Less 'not Q'	More 'not Q'
If P then NOT Q	Less 'not Q'	More 'not Q'	More 'Q'	Less 'Q'
If NOT P then Q	More 'Q'	Less 'Q'	Less 'not Q'	More 'not Q'
If NOT P then NOT Q	Less 'not Q'	More 'not Q'	More 'Q'	Less 'Q'

As pointed out previously, the materials used in the present study do not manipulate the concept sizes represented on the cards within the task. The mental activation model does not specify that the cards will be re-coded into representations of different sized sets, and therefore differences in selection rates as in the above Table 4.1 or as Yama found between tasks are not predicted. If re-coding is not taking place, then just the concepts on the cards will be activated in memory and the normal pattern of matching bias will be observed. The present study attempts to control for the concept sizes represented by the cards and only manipulates the sizes of the concepts mentioned in the premise. Thus, the two antecedent and two consequent cards represent roughly equal sized concepts.

Predictions

Matching Bias within Tasks

Up until now, there has been no specific definition in reasoning problems as to what defines 'matching'. Cases could match along several dimensions, two of which are *lexically identical* or *semantically related* in some way. All accounts of matching bias would predict a general matching response in the lexical condition, as it mirrors the standard abstract form of the selection task. The HA and mental model theories contain no specific predictions with regard to the lexically mismatching materials used in the semantic condition. They offer no details as to how the information contained within a problem is processed, only that matching items may be perceived as more relevant than others. The ODS account suggests that the probability of occurrence of the premise materials cause apparent biasing effects. The items on the cards are therefore processed with respect to the items mentioned in the premise so matching bias would be predicted by the ODS account to occur in both conditions. The mental activation model also predicts matching in both conditions, because it is assumed that the concepts mentioned in a problem activate a distributed memory network that stores concepts at both a lexical and semantic level.

Matching Bias between Tasks

Although differences in perceptions of relevance between lexical and semantic conditions should be taking place according to the mental activation model, it is unlikely that the selection task will reveal them. The selection task only allows differences in relevance judgements of the cards within the task. Although the 'oak' card in the semantic condition might be seen as less relevant than the 'oak' card in the lexically matching condition, it is likely that the 'oak' card in both conditions will still be seen as more relevant than the 'ivy' card and is equally likely to be chosen in both conditions. Although Yama (2001) found differences in card selection rates between tasks it is argued that a greater effect in those studies was created using bounded contrast sets that facilitated the re-coding of the cards into different sized concepts.

If the selection task were sensitive enough to detect differences in perceptions of relevance between tasks, then there may be found to be differences between conditions in the present study. The activation model would predict a pattern of more overall matching card selections in the lexical task than the semantic task but it is unlikely to be significant. In the lexical task the items completely overlap in memory representation and might thus appear as more relevant. The ODS account would also predict this pattern of more matching 'Q' cards in the lexical condition than in the semantic condition. The ODS predictions concerning the probability of 'Q' also state that with a high probability 'Q' (semantic condition) there will be more 'not Q' selections, than with a low probability 'Q' (lexical condition). The mental activation model would not account for this difference.

Repetition & Semantic Priming

If simple priming effects were responsible for matching bias then the two conditions may be seen to elicit differences in matching bias. Repetition priming is generally more robust than semantic priming and if it were responsible for matching bias then the matching bias effect would only be seen in the lexical condition. If both priming effects were responsible, then there may be a contrast in matching bias effect size, with the lexical condition again eliciting a greater amount, due to the materials in the task being of a both lexical and semantic nature.

Lexical Matching Strategy

Oaksford (2002) claimed that when explicit negation is used, reasoners could possibly be adopting a lexical matching strategy (*matching*$_2$). It was pointed out in Chapter 1 that

Oaksford believes $matching_2$ occurs with the whole clause in the premise. If the clause 'not a 2' occurs in the premise, an explicitly negated card which also displayed 'not a 2' would match and would therefore be more likely to be chosen. Such a phenomenon is offered as an explanation for differences found using implicit and explicit negations within a reasoning task. If reasoners adopted this lexical matching strategy in reasoning tasks, then it would be unusual for this to only occur when explicit negations were used. If the lexical $matching_2$ strategy was being used by a large proportion of reasoners then it should be expected that such a strategy were also being employed in the standard task using implicit negations. Therefore, it could be expected that there would be more $matching_2$ taking place in the lexical condition.

Experiment 1

Participants

A total of 99 undergraduate students at the University of Plymouth participated either as paid volunteers or as part of their course requirement. The experiment took place in two sessions, with participants being randomly allocated to one of the two conditions within each session (The total for each condition being 50 participants in the 'lexical' condition and 49 participants in the 'semantic' condition).

Design and Procedure

Participants were tested in two groups: lexically matching premise items and semantically matching premise items. In each condition, participants would receive a booklet containing four selection task problems. The selection task cards were identical in each condition, the difference between the two conditions being that in the semantic group, participants would be given inference items that represented the superordinate category of the items on the cards. For example, having four cards with the words, 'COD', 'FOX', 'OAK' and 'IVY', in the semantic condition, participants would be given a premise, "If a card does NOT have a *fish* on the left side, then it has a *tree* on the right side". For the lexical condition the premise might have read, "If a card does NOT have a *cod* on the left side, then it has an *oak* on the right". Thus, the two premise items either semantically match two of the four cards or they lexically match. The entire card contents and appropriate superordinate categories can be found in Appendix B.

The use of negations was controlled for by the use of the 'negations paradigm'. Here, negations in the premise are controlled for, constructing the following four premises by systematically altering the position of the negations within the sentence:

1. If a card has a P on the left side, then it has a Q on the right side.
2. If a card has a P on the left side, then it does <u>not</u> have a Q on the right side.
3. If a card does <u>not</u> have a P on the left side, then it has a Q on the right side.
4. If a card does <u>not</u> have a P on the left side, then it does <u>not</u> have a Q on the right side.

The four sets of materials used for the problems in each condition were allocated to one of these four premise statements by means of a Latin square. Each participant received all four negations paradigm formats with materials systematically allocated to each premise type. For example, the materials used in the above example could have been allocated to either of the four negations paradigm premises.

Instructions

Each booklet of four problems had the following instructions on the cover page:

> *"Attached you will find four problems – one on each sheet. Each problem consists of a rule followed by pictures of four cards. Your task is to select <u>any</u> or <u>all</u> of the cards that you think would be necessary to provide enough information to discover whether the rule given is true or not.*

All cards consist of two halves – a right side and a left side. You will only see information on one of the sides – the other side will be hidden from view. It is your task to select the cards whose 'hidden' information could test the rule once revealed.

Please think carefully about your selections and circle the cards that you have chosen."

Results

The percentage of card selections for both the lexical and semantic conditions can be seen in Table 4.2. Where the cards matched, that is where they represented an affirmative form of the premise item, they have been underlined. Matching bias is usually seen when for each logical case (i.e. TA, FA, TC, and FC), the two rules where the card matches produce higher selection frequencies than the two that do not match. Thus it can be seen that this matching bias trend was present in both the lexical and semantic conditions.

Table 4.2. Percentage of Cards Selected in Experiment 1 by Logical Case.

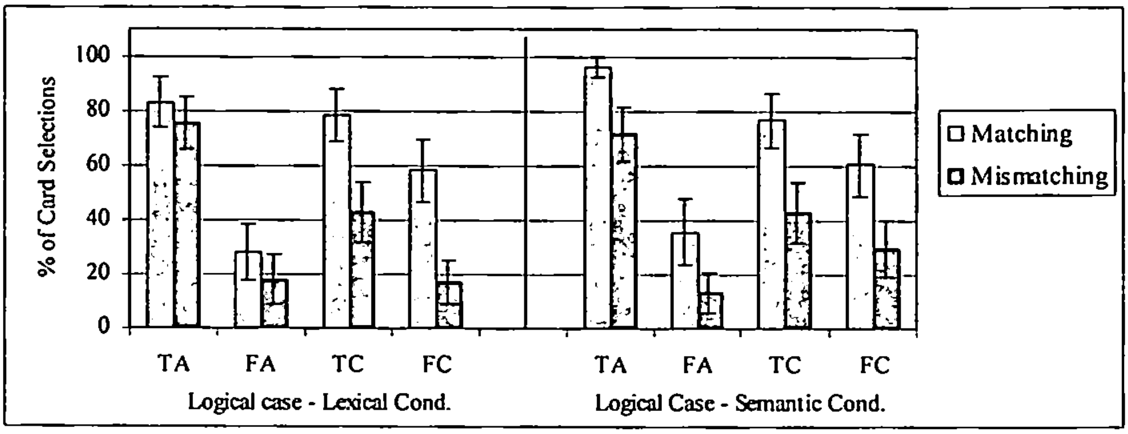
Lexical Condition	TA	FA	TC	FC
If P then Q	88	18	74	18
If P then NOT Q	78	18	34	66
If NOT P then Q	72	30	82	16
If NOT P then NOT Q	78	26	52	50

Semantic Condition	TA	FA	TC	FC
If P then Q	96	18	76	29
If P then NOT Q	96	8	37	67
If NOT P then Q	73	27	78	31
If NOT P then NOT Q	69	45	49	53

NB. Matching cases underlined

If the matching instances are combined for each logical case they can be compared to the two mismatching cases (also combined) giving a clearer overview of the matching taking place for each logical type (Figure 4.2).

<u>**Figure 4-2.**</u> The Percentage of Matching and Mismatching Card Selections by Logical Case (Exp.1).



(NB Error bars indicate +/- 2 standard errors)

It can be seen from Figure 4.2 that there are more matching cards than mismatching cards selected for each logical case and there are overall more TA than FA responses and more TC than FC responses (confirmation bias). Thus the two conditions have produced card selection frequencies typical of previous findings in the literature with regard to implicit negations (see Figure 4.1 – Evans et al, 1996). Matching bias is present throughout, and there is also a general pattern of confirmation bias.

Matching indices can be formulated in order to see if matching bias is present. Following Evans, Clibbens and Rood (Exp.2.,1996), matching indices are divided into *antecedent matching indices* (AMI) and *consequent matching indices* (CMI). These are simply calculated by counting the number of matching TA and FA card selections and subtracting the amount of mismatching TA and FA card selections to give the antecedent matching indices. The consequent matching indices are calculated similarly for the consequent card selections (i.e. TC and FC). The *logic index* (LI) can also be calculated to give an indicator as to the amount of logically correct selections that may be taking place. This is calculated in line with Pollard and Evans (1987) who added a one for every logically correct (TA or

FC) card selected and subtracted a one for every logically incorrect card selected (FA and TC). All indices are calculated across all four problems and therefore can range from -4 to +4 in the case of the matching indices (which are each only calculated for half of the cards) and from -8 to +8 in the case of the logic indices.

One-sample t-tests were carried out to establish whether the index scores were significantly different from zero (null hypotheses) for each group – an indicator that the matching bias effect is occurring. If matching bias were not present then the indices should score close to zero. The following Table 4.3, depicts the indices scores and their statistical significance.

<u>**Table 4.3.**</u> T-test Values for the Antecedent Matching, Consequent Matching and Logic Indices (Exp.1).

Condition	AMI	CMI	LI
Lexical[a]	1.88*	6.69*	4.08*
Semantic[a]	5.26*	5.02*	4.92*
t' - between[b]	-2.21*	0.68	-0.94
	(*df*=96)	(*df*=95)	(*df*=95)

*denotes significance
[a] *one-tailed* test for significance (P<.05) from zero
[b] *two-tailed* test for significance (P<.05) between conditions

In all cases, the matching indices were significantly different from zero using a one-tailed, one-sample t-test. Therefore, in these cases, the matching bias effect could be seen to be very much in existence. The results also show that there was a significant matching effect difference between conditions for the antecedent cards (AMI scores). It would appear that there was a significantly greater matching effect taking place with semantic materials than with lexical materials for antecedent card choices. From Figure 4.2 it appears that the difference in matching indices for antecedent cards between the lexical and semantic conditions is mainly due to more matching antecedent cards being selected in the semantic condition (Approx. 13% more matching TA cards). Both conditions seem to elicit similar levels of mismatching antecedent card choices. There was also a significant amount of logically correct responses indicated by the logic indices for both conditions.

As there was found to be no significant difference between the CMI scores, the significant AMI difference at present may be due to some anomaly with the materials used for the antecedent cards. Thus Experiment 2 sets out to clarify the present situation by reversing the materials used for the antecedent and consequent clauses.

Experiment 2

The results from Experiment 1 showed unexplained differences between the antecedent card selections and consequent card selections. If there were differences between the lexical and semantic conditions, this would be expected to affect both the AMI and CMI scores. These results may have occurred because of the placement of materials within the problem context. It was therefore necessary to replicate the study but this time reversing the materials previously used to represent the antecedent and consequent card selections. Therefore both the antecedent/consequent was changed in the premise and also on the cards. The aim of this experiment was to see if the results in Experiment 1 were due to any peculiar interaction of the materials or whether there was a genuine effect in terms of the differences found previously between the antecedent matching indices.

Participants

A total of 100 undergraduate students at the University of Plymouth participated either as paid volunteers or as part of their course requirement. The experiment took place over the course of one week, with participants being randomly allocated to one of the two conditions. Participants were tested in groups of up to six. The total for each condition being 48 participants in the *lexical* condition and 52 participants in the *semantic* condition.

Design and Procedure

Participants were again tested in two groups as in Experiment 1: lexically matching premise items and semantically matching premise items. In each condition, participants would receive a booklet containing four selection task type problems. As before, the selection task cards were identical in each condition, the difference between the two conditions again being that in the semantic group, participants would be given inference items that represented the superordinate category of the items on the cards. This time the materials were reversed, in that materials used for the antecedents in Experiment 1 were now being used for the consequent materials and likewise for the Experiment 1 consequent materials. The instructions and general format were the same as in Experiment 1.

Results

The matching and logic indices were again calculated and can be seen in Table 4.4. It would seem that the AMI scores for the semantic condition are very much smaller than in Experiment 1. The lexical AMI scores are no longer significantly above zero whilst the CMI scores are only just significant ($P = 0.47$). There is no longer a significant difference between AMI scores for both conditions. It must be assumed therefore that the significant effect found in Experiment 1 was an anomaly caused by the participant population used at that time and could not be replicated. If the AMI difference in Experiment 1 was due to the materials, then this should now have produced a significant difference in CMI scores, which was not apparent in this second experiment.

Table 4.4. T-test Values for the Antecedent Matching, Consequent Matching and Logic Indices (Exp.2).

Condition	AMI	CMI	LI
Lexical[a]	1.65	5.33*	5.03*
Semantic[a]	1.71*	4.87*	7.40*
t' – between[b]	0.28 (*df*=89)	0.89 (*df*=94)	0.33 (*df*=79)

*denotes significance
[a] *one-tailed* test for significance (P<.05) from zero
[b] *two-tailed* test for significance (P<.05) between conditions

The combined analysis for Experiments 1 and 2 will now be addressed and will be used for discussion purposes.

Combined Results

Table 4.5 shows the percentage of cards selected combining both experiments. There are still more matching cards selected (underlined) within each logical case, supporting the matching prediction

Table 4.5. Percentage of Cards Selected (Experiments 1 & 2 Combined) by Logical Case.

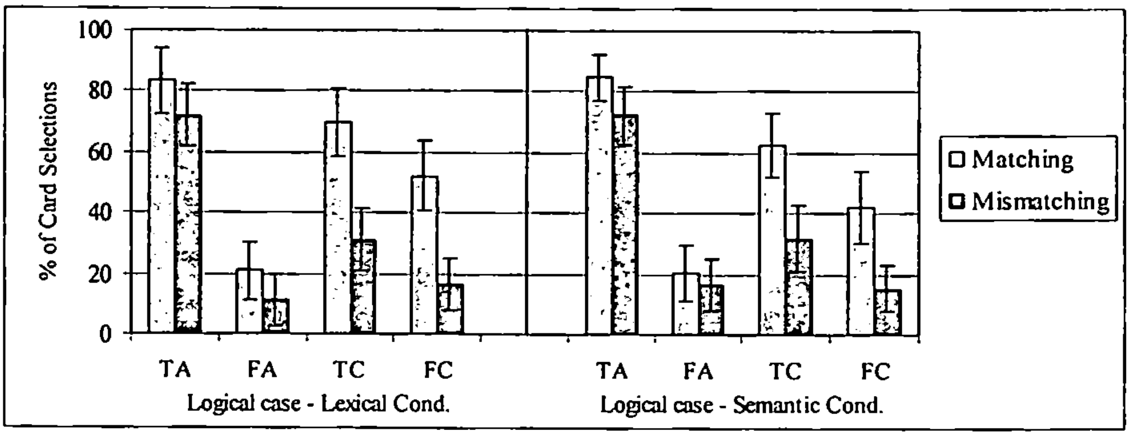
Lexical Condition (n=98)	TA	FA	TC	FC
If P then Q	<u>86</u>	14	<u>70</u>	16
If P then NOT Q	<u>81</u>	15	27	<u>61</u>
If NOT P then Q	73	<u>22</u>	<u>78</u>	17
If NOT P then NOT Q	73	<u>27</u>	48	<u>49</u>

Semantic Condition (n=101)	TA	FA	TC	FC
If P then Q	<u>88</u>	18	<u>66</u>	24
If P then NOT Q	<u>92</u>	12	33	<u>52</u>
If NOT P then Q	70	<u>22</u>	<u>72</u>	21
If NOT P then NOT Q	73	<u>34</u>	42	<u>50</u>

NB. Matching cases underlined

It can be seen from Figure 4.3 that a general matching bias phenomenon is occurring within all four logical cases across both the lexical and semantic conditions. There are also more TA and TC cards selected, revealing a confirmation bias or bi-conditional reading of the premise.

Figure 4-3. The Percentage of Matching and Mismatching Card Selections by Logical Case (Exp.2).



(NB Error bars indicate +/- 2 standard errors)

The matching and logic indices were calculated as before and show significant matching effects and a significant amount of logical responses (TA & FC). With the data from the two experiments combined, there is seen to be no significant difference between the two conditions (Table 4.6) with regard to matching or logic.

Table 4.6. T-test Values for the Antecedent Matching, Consequent Matching and Logic Indices (Exps. 1 & 2 combined).

Condition	AMI	CMI	LI
Lexical[a]	2.48 *	8.44 *	6.45 *
Semantic[a]	4.65 *	7.03 *	8.23 *
t' - between[b]	-1.14	1.14	-0.51
	(*df*=191)	(*df*=196)	(*df*=192)

*denotes significance
[a] *one-tailed* test for significance (P<.05) from zero
[b] *two-tailed* test for significance (P<.05) between conditions

With regard to the possibility that reasoners might choose more 'not Q' cards in the semantic condition (high probability Q) and more 'Q' cards in the lexical condition (low probability Q) a 2x2 mixed ANOVA was carried out. The two factors were condition (lexical/semantic - *between*) and consequent card selection (Q/not Q- *within*) with the data being collapsed across logical case. Following the predictions of Oaksford, Chater and Grainger (1999), there should be a significant interaction between the two factors, such that there would be more 'Q' card selections in the lexical condition and more 'not Q' card selections in the semantic condition. The analysis showed that there was no significant interaction ($F_{1,197}$ = 1.292, MSE = 1.532, p = .257) although the general pattern of responses was in the predicted direction.

If the predictions from Table 4.1 are compared with the mean selection frequencies in Table 4.5, it can be seen that the general pattern follows the ODS predictions when separated by logical case (TC & FC cases). On all four negations-paradigm inferences there are more 'Q' card selections in the lexical condition than in the semantic condition. In the semantic condition, three-out-of-four negations-paradigm inferences elicited more 'not Q' selections than in the lexical condition.

Discussion

The use of superordinate categories in the semantic premises has not changed the matching bias phenomenon. The same matching effect is taking place using 'mismatching' premise clauses and cards as there is found using identical words in the lexical condition. The matching bias effect is therefore shown to be independent of lexically matching terms. Matching in the selection task may be defined as a tendency to choose a card that reflects the topic item in the premise irrespective of whether it lexically matches or not. The matching effect up until now has been poorly defined, as is highlighted by Yama's (2001) experimental replication (Exp. 3) of his first experiment, using lexically matching materials instead of the original semantically matching ones (Exp. 1). It was obviously felt that matching bias could only be elicited when the materials lexically matched.

It would also appear that simple repetition priming effects, which would have shown differences in matching between conditions, do not cause matching bias. This is not surprising, as the selection task format differs from the task format used in repetition priming studies. In *repetition* priming studies, the participants usually cannot remember the items from the initial priming task. The subsequent effects on performance in a task are very much at an unconscious level. With *semantic* priming, again the prime is seen before the target item and affects performance. In the selection task, both prime and target as it were (premise clause and matching card), are seen together simultaneously and for a duration that is controlled by the reasoner, not the experimenter. They can re-read the problem items as many times as they wish and for as long a period as they require in order to make a response.

This is not to say though, that the underlying mechanisms for priming are not also responsible for matching bias. The mental activation model is based on the same incremental learning approach to memory that is also offered as an explanation for priming effects. The incremental learning effects proposed to stem from the problem materials can account for the greater and more durable effects than those found in priming tasks. Especially in semantic priming tasks, the performance effects are shorter than repetition priming effects. In a reasoning problem like the selection task, the repeated and extended exposure to the stimuli could cause a greater degree of incremental learning in the connections between nodes, than that found in priming studies. Repeated activation of the same or similar concepts in memory facilitates subsequent activation of those concepts. It has been argued that this facilitation in activation can account for the heightened relevance or focussing effects described by the HA and mental model theories. The heightened perceived relevance of some of the items in the problem (matching ones) creates a selective mental representation of the problem and thus leads to biased card selections.

The study also sheds some light on the issue of explicit and implicit negation. There was shown to be a significant matching effect using both lexically matching and semantically related materials. It would seem unlikely that greater responses to explicitly negated cards (as in Evans, Clibbens & Rood, 1996) are, as Oaksford (2002) claims, possibly due to a $matching_2$ strategy whereby reasoners are matching the identical clauses e.g. 'not a 2' in the premise with 'not a 2' on the card. It seems reasonable to assume that if participants in such tasks were matching_2 the identical premise clause *with negation* then they would be likely to adopt a similar matching_2 strategy with regard to the affirmative clause *without negation* as in the present study ('oak' in the premise, and 'oak' on the card). In which case, there should be found to be more matching selections in the *lexical* condition where the word in the premise is identical to the word on the card. In the *semantic* condition there should be less matching_2 as no premise clause (with or without negation) identically matches the cards. This was not the case.

The results from this study pose a further question concerning the use of implicit/explicit negation. Both types of negation are *semantically similar* in the context of the problem. Given the consequent clause '...then there is a not a 2 on the other side' the implicitly negated card ('3') and explicitly negated card ('not a 2') both referentially match the premise clause. If the matching effect occurs with semantic materials when they *represent* each other but do not lexically match, why are more responses found to be given to explicitly negated cards as opposed to implicitly negated cards when they both represent the same logical referent (Evans, Clibbens & Rood, 1996)? The ODS theory does not account for explicit/implicit differences as it is based on the probabilities represented by the cards. According to the ODS theory, the implicitly negated '3' card still represents 'not a 2'. Therefore, there should be no differences between explicit and implicit tasks.

The matching bias results from this study show that there is still a selection preference for the affirmative rather than implicitly negated cards in both conditions. It would seem that semantic association between the cards and the premise clauses is stronger for affirmative clauses than negated ones. In other words, reasoners must have greater difficulty in seeing the semantic association between the explicitly negated premise clause (e.g. 'not an oak') and the implicitly negated cards (e.g. 'ivy'). So, although matching is still taking place at a semantic level, it would appear that the semantics of negation (within the logical context of the problem) are not being seen to be as relevant. Where for instance an 'oak' is semantically interpreted as a 'tree', 'ivy' is less likely to be semantically interpreted as 'not a tree'.

Evans (1998) describes how the presence of negations linguistically cue the topic of the sentence which explains why matching bias is shown as a preference for the affirmative card irrespective of negation in the premise. Evans and Handley (1999) subsequently explained matching bias as a tendency to see mismatching items as irrelevant. This study shows that problem items can mismatch lexically but can produce the matching effect semantically. The notion of implicit cards being seen as *irrelevant* can only be because either the semantic interpretation of the implicit negation isn't being made or there are representational anomalies in terms of memory. The latter account seems more favourable in the light of the present results. If semantic interpretations of affirmative instances can be made just as easily as if the items lexically matched, then why can't the semantics associated with implicit negation be interpreted just as easily? In the semantic condition, both consequent cards (affirmative & negative) mismatch lexically but can both match semantically with the premise clause (affirmative & negative) if they are being interpreted in terms of the problem. For instance, the 'oak' card semantically can match the premise clause 'tree', and the 'ivy' card can semantically match the premise clause 'not a tree'. Therefore, Evans and Handley's account of *mismatching* remains undefined.

The mental activation account describes how the semantic association between the premise clause and the affirmative card is made easily, but the semantic association of the implicitly negated card to the negated premise is not made. The mental activation account describes how two semantically related concepts can cause overlapping activation areas. The two items are not so much semantically interpreted with respect to the problem but they are *perceived* to be related by the way memory is organised. The implicit negation on the other hand is an unrelated concept and therefore would not produce an activation overlap with the activation associated with the premise item. It will be remembered that only explicit negation can be represented in the activation model, either by a reduction in activation for the affirmative concept, or by the activation of the affirmative concept together with the activation of some related concepts that might represent the negation. An implicit negation in a reasoning problem is not interpreted as such, but is represented in its affirmative form.

Whilst the ODS theory relies heavily on the implicit negation being reinterpreted as representing a contrast set and therefore a larger probability of occurrence, in doing so it can not explain differences in implicit/explicit response patterns. If the ODS theory is amended with the notion of matching_2, the findings concerning lexical matching from the present study cause the theory further difficulties. As was pointed out earlier, there is no convincing reason to assume that lexical matching takes place only with explicitly negated items. If lexical matching were taking place, then the matching effect would be greater in the lexical condition in this study.

Although the ODS theory would predict differences between tasks as were outlined in Table 4.1 these were not significant. It was suggested though that the selection task may not be sensitive enough to detect differences between conditions. Both the ODS theory and the mental activation model would predict differences in response patterns for matching 'Q' cards – with the lexical condition eliciting generally more 'Q' card selections than the semantic condition, which was observed in this study. The ODS account and mental activation model differ with regard to 'not Q' selection pattern predictions. The mental activation model predicts that there would be more 'not Q' selections in the lexical condition than in the semantic condition because of the smaller areas of activation that would be involved in representing the smaller concepts. The ODS account claims that there would be more 'not Q' selections in the semantic condition because 'Q' in this instance represents a higher probability of occurrence. Three out of the four negations paradigm inferences in the semantic condition elicited more 'not Q' card selections. Such a pattern of responses therefore offers more support to the ODS account than to the mental activation account.

At present, the activation model seems to predict matching bias and relevance perceptions with more specificity than either the current HA or mental model accounts. It offers a working definition of *mismatching*, not specifically in terms of lexical matching or semantic interpretations but in terms of memory activation overlap of the items mentioned in the problem. It also accounts for implicit negation without reference to interpreting contrast sets and therefore does not encounter difficulties with explaining explicit negation effects.

Experiment 3

Introduction

In Experiments 1 & 2 it was found that the matching bias effect could still occur in the selection task with semantically related materials. It was seen that the items mentioned on the cards did not have to be lexically identical to the concepts mentioned in the premise. It was sufficient to cause the matching bias effect to have cards that reflect the premise items by way of a semantic relationship. It appeared that this semantic similarity between the card and the premise item in some way made that card appear more relevant to the task than alternative cards which did not share this semantic relationship.

The proposed mental activation model (see Chapter 3) accounted for this disparity in perceived relevance of the cards by means of describing the activation in memory of all the concepts mentioned in the selection task problem (two concepts in the premise and four concepts on the cards). Only those cards, which could be classed as being semantically related to the items mentioned in the premise, would overlap in memory activation. This activation overlap was proposed to heighten or facilitate levels of overall activation, causing those cards to appear as more relevant. This heightened perception of relevance might thus cause biased responses when reasoners were uncertain as to which cards to select. In a way, this biased selection process could be seen as an unconscious form of intuition and the mental activation overlap is not dependent on lexically matching items so long as the concepts initiate activation of the same or similar distributed network.

This final selection task experiment sets out to further test the proposed mental activation model. As was previously pointed out, the mental activation model suggests that matching bias in the selection task is caused by the fact that only the matching cards mentally activate similar parts of the semantic network that are activated by the premise items. For instance, if we call the two premise items that represent superordinate categories 'A' and 'B', and the cards that represent category instances as 'a', 'x', 'b' and 'y', the 'a' and 'b' cards activate part of the same semantic network as the premise items. Cards 'x' and 'y' do not activate parts of the semantic networks that are activated by the premise items and therefore do not 'overlap' in activation. They are therefore perceived as less relevant to the task as their activation levels have not been heightened due to overlapping activation. If on the other hand, cards 'x' and 'y' were also semantically related to the two premise items, then according to the mental activation model, all four cards would appear as equally relevant – they would all overlap in activation to the two premise items and matching bias would diminish.

To create a selection task in which alternative (by alternative it is meant that the cards do not match their respective antecedent and consequent premise items – *false antecedent & false consequent*) card choices still semantically overlap the premise items, the same task format can be used as in Experiments 1 and 2. Obviously, having all four cards that semantically represent the premise items necessitates two of the cards being presented on alternate sides of the cards. For instance, given the premise, 'If a card has a fruit on the left side, then it has a vegetable on the right side' would instigate having two items that represent fruit ('apple' & 'orange') and two items that represent vegetable ('potato' & 'carrot'). There would be no negating cards if both fruits appeared on the left side of the cards and both vegetables on the right. All four cards would represent true antecedent (TA) and true consequent (TC) cases. So to circumvent this, one 'fruit' card could represent the TA by being placed on the left side of the card, whereas the other 'fruit' could represent the FC by being placed on the right side of the card. Likewise the two 'vegetables' could be placed in the same positions respectively to represent the FA and TC options. Thus, a selection task can be created in which all the card items effectively match the premise items in terms of overall mental activation. The only difference from standard matching cards is that two of the semantically related cards are in different positions and represent different aspects of the problem.

It was suggested in Experiments 1 & 2 that card selections alone may not offer a detailed view of the relevance judgements that are taking place in the task. A selected card offers no indication of the perceived relevance of the non-selected cards. Also, other task factors such as the logical requirement demanded by the problem may lead reasoners to select cards on another basis away from the perceived relevance that may be created by the materials. It was therefore necessary to create a task in which not only were cards selected but also where all cards were rated on scale. The scale that was devised measured participants' perceptions of the *likelihood* that a particular card could falsify the premise. It was felt that such a scale would reflect participants' perceptions of the relevance of each card in relation to the task.

Such a scale appears to not have been used in any selection task studies to date. The only card judgements that have appeared in a selection task were instigated by Oaksford, Chater and Grainger (1999) who used a Probability Rating Scale (PRT). The PRT involved asking participants, either before or after the selection task, about their judgements concerning the probability of occurrence of the items mentioned in the premise e.g. 'Of every 100 people, how many would you expect to be politicians'? The PRT also asked for probability judgements concerning the co-occurrence of the two premise items e.g. on a scale of 0-100%, what is the likelihood that, 'If a person is a politician, then they are privately educated'? The PRT made no enquiries concerning the alternative (FA & FC) card concepts that were displayed.

No other theory of reasoning biases has concerned itself with these alternative or mismatching card options. Likewise, none of these theories to date would have an explanation for any differences in responses caused by manipulating the similarity/dissimilarity of these alternative cards. Only the proposed mental activation model predicts that a reduced level of perceived relevance would occur to matching cards using all semantically similar items.

Participants

Sixty-three undergraduate psychology students at the University of Plymouth participated in the experiment, either as part of their course requirement or for payment. Participants were tested in small groups of up to six at a time. Participants were randomly allocated to either of two groups: 32 participants in the *similar* condition and 31 participants in the *dissimilar* condition.

Design and Procedure

Participants were tested in two groups: semantically *similar* alternative card choices and semantically *dissimilar* alternative card choices. In each condition, participants would receive a booklet containing four selection task problems. The conditional premises and subsequent 'matching cards' were identical in both conditions. The two conditions differed in the content of the 'mismatching cards'. In the *similar* condition, the mismatching (alternative) cards were constructed so that they were semantically similar to the concepts mentioned in the premise. For instance, given the premise, 'If a card has a fruit on the left side, then it does not have a vegetable on the right side', the four cards depicted 'apple' (left), 'potato' (left), 'carrot' (right) and 'orange' (right). The four cards were presented so that a word appeared on either the right or left side, this being shown here in parentheses. Using the above example, it can be seen that two types of fruit and two types of vegetable were displayed on the cards. So that the two semantically related items could be used within the logical structure of the problem, they were placed on two different sides of the cards (right & left). In the above example, the two fruits were used to represent 'P' and 'not Q' and by doing so the 'apple' card *matches* the antecedent concept of 'fruit' and the 'orange' card *mismatches* the consequent item of 'vegetable'. In the *dissimilar* condition, items that are semantically unrelated were used for the mismatching cards i.e. given the premise, 'If a card has a fruit on the left side, then it does not have a vegetable on the right side', the cards on display were 'apple' (left), 'glass' (left), 'carrot' (right) and 'fox' (right). The two items, 'glass' and 'fox' are semantically unrelated to either of the two premise items ('fruit and 'vegetable'). The full list of the items used in both conditions can be seen in Appendix C together with an example of the problem layout.

In addition to the standard task of card selection, participants were required to mark on a five-point scale (*1 = unlikely, 5 = likely*) how likely/unlikely they thought each particular card could falsify the premise if that card were unmasked and the item on the other side revealed.

The use of negations was controlled for by the use of the 'negations paradigm'. Here, negations in the premise are controlled for by systematically altering the position of the negations within the sentence:

1. If a card has a P on the left side, then it has a Q on the right side.
2. If a card has a P on the left side, then it does <u>not</u> have a Q on the right side.
3. If a card does <u>not</u> have a P on the left side, then it has a Q on the right side.
4. If a card does <u>not</u> have a P on the left side, then it does <u>not</u> have a Q on the right side.

The four sets of materials used for the problems in each condition were allocated to one of these four premise statements by means of a Latin square. Each participant received all four negations paradigm formats with materials systematically allocated to each premise type. For example, the materials used in the above example could have been allocated to either of the four negations paradigm premises.

Instructions

Each booklet of four problems had the following instructions on the cover page:

> *"Attached you will find four problems – one on each sheet. Each problem consists of a rule followed by pictures of four cards.*
>
> *All cards consist of two halves – a right side and a left side. You will only see information on one of the sides – the other side will be masked (hidden) from view.*
>
> *Your Task:*
>
> *For each card, mark on the scale how* ***likely*** *you think that particular card could falsify the rule if the mask was taken off the other side, and the other item revealed.*
>
> *Secondly,* ***circle*** *the card(s) that you would unmask in order to see if the rule is true or false.*
> *Please think carefully about your selections.*

Results

Card Selection Rates

The percentage of selections for each card in both the similar and dissimilar conditions can be seen in Table 4.7. Where the cards matched in respect to the corresponding item in the premise, they have been underlined. (NB although all cards in the *similar* condition matched either of the two premise items, they would only be considered *matching* if they matched the respective antecedent or consequent clause by way of their position on the cards i.e. antecedent items were always on the left side of the cards). As previously in Experiments 1 and 2, matching bias can be seen when for each logical case, the two cards that match produce higher selection frequencies than the cards that do not match. It can be seen therefore that matching bias was very much in existence in both conditions.

Table 4.7. Percentage of Cards Selected in Experiment 3 by Logical Case.

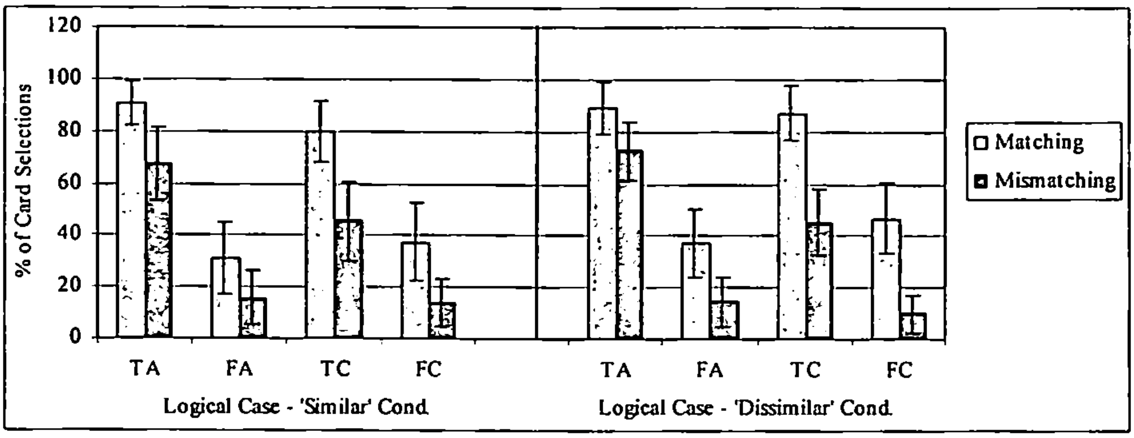
Similar Group	TA	FA	TC	FC
If P then Q	91	16	84	9
If P then NOT Q	91	16	44	38
If NOT P then Q	63	28	75	19
If NOT P then NOT Q	72	34	47	38

Dissimilar Group	TA	FA	TC	FC
If P then Q	90	10	81	10
If P then NOT Q	87	19	29	55
If NOT P then Q	71	42	94	10
If NOT P then NOT Q	74	32	61	39

NB. Matching cases underlined

To give a better picture of the matching bias effect, all matching and all mismatching instances are combined for each logical case as can be seen in Figure 4.4. It is quite apparent that similar levels of selections and matching bias occurred in both conditions. The card selection frequencies for both conditions are therefore typical of previous selection task studies (see Exps.1 & 2). It would seem therefore that the similarity manipulation has not affected the general pattern of card selections.

Figure 4-4. The Percentage of Matching and Mismatching Card Selections by Logical Case (Exp.3).



(NB Error bars indicate +/- 2 standard errors)

A more accurate measure of any matching bias effects can be ascertained by calculating the antecedent and consequent matching indices (AMI & CMI). As before, these are simply calculated by counting the number of matching TA and FA card selections and subtracting the number of mismatching TA and FA card selections (AMI) and likewise for the TC and FC card selections (CMI). The logic indices (LI) are also calculated in order to indicate the amount of correct selections that are taking place. A one-sample t-test was carried out on all these indices to indicate whether a significant amount of these selection types was taking place (indicated if the indices are significantly different from zero). The results can be seen in Table 4.8.

Table 4.8. T-test Values for the Antecedent Matching, Consequent Matching and Logic Indices (Exp.3).

	AMI	CMI	LI
Similar [a]	2.49 *	3.71 *	2.21 *
Dissimilar [a]	3.36 *	5.87 *	2.10 *
t' – between [b]	0.02	-1.03	0.15
	(*df*=61)	(*df*=61)	(*df*=61)

*denotes significance
[a] *one-tailed* test for significance (P<.05) from zero
[b] *one-tailed* test for significance (P<.05) between conditions

The t-test results confirm that there is a significant amount of matching bias taking place in both conditions. It would appear that the similarity manipulation has had no significant affect on the amount of matching and mismatching cards that participants selected. Neither has the experimental manipulation affected the amount of correct responses between groups, who both show significant selections of logically correct cards as depicted by the logic indices. It was stated in the introduction that the amount of cards selected might not be a precise enough measure of participants' perceptions of each card's relevance. What may give an overall clearer picture of relevance judgements are the ratings participants give for the likelihood of each card being able to refute the given rule. The rating scale results will now be looked at.

Rating Scale

Participants' ratings on the five-point scale were re-coded from 1-5 (unlikely-likely) to –2 to 2 respectively. This meant that a minus score indicated that participants thought 'unmasking' that particular card was unlikely to provide information that could falsify the rule. In the same manner, a positive score would indicate that the card could falsify the rule. The further the rating score was from zero, the more confident the participant was concerning the likelihood of the card falsifying the rule. Table 4.9 shows the average rating for each card. It appears that the *dissimilar* condition elicited ratings that were the furthest from zero. It would appear that using dissimilar cards made reasoners more confident in their judgements of whether the cards were likely or unlikely to falsify the premise.

Table 4.9. Average Rating for each Card in Experiment 3 by Logical Case.

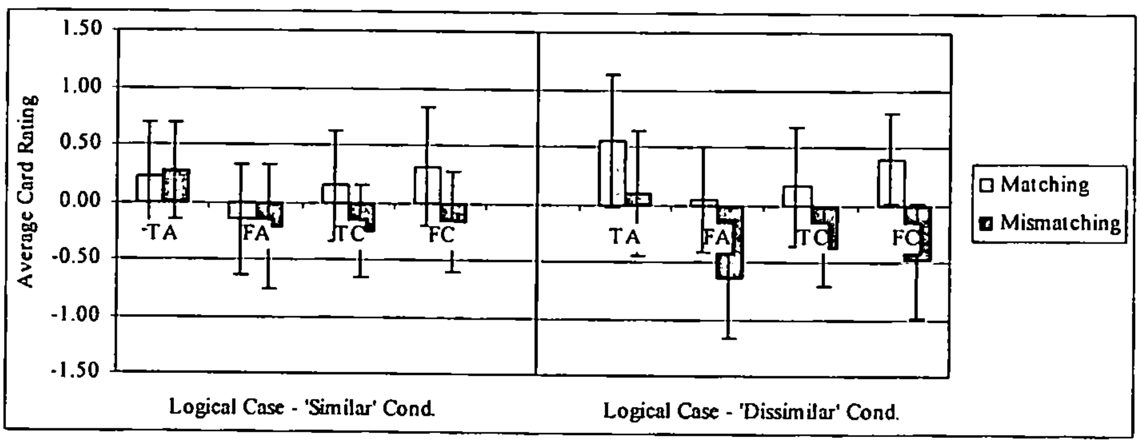
Similar Group	TA	FA	TC	FC
If P then Q	0.05	-0.02	0.02	-0.05
If P then NOT Q	0.39	-0.44	-0.23	0.30
If NOT P then Q	0.14	-0.20	0.27	-0.30
If NOT P then NOT Q	0.39	-0.11	-0.27	0.33

Dissimilar Group	TA	FA	TC	FC
If P then Q	0.53	-0.42	0.08	-0.42
If P then NOT Q	0.56	-0.89	-0.65	0.53
If NOT P then Q	0.11	-0.15	0.24	-0.56
If NOT P then NOT Q	0.08	0.24	-0.11	0.27

NB. Matching cases underlined

Figure 4.5 shows the collapsed average scores for matching and mismatching cards in each logical case. A minus score indicates the degree that a card is perceived as unlikely to falsify. It is quite apparent that the *dissimilar* condition created the largest variation of ratings scores. When cards matched the premise item, reasoners were far more inclined to rate the cards as *likely* to falsify the rule. When the cards mismatched, they were far more inclined to rate the cards as *unlikely* to falsify the rule. It would appear that manipulating the semantic relationship of the mismatching cards has not only affected their perceived likeliness, but has also affected participants' perception of the matching cards. It will be remembered that the matching cards were identical in each condition.

Figure 4-5. The Total Ratings for Matching and Mismatching Cards by Logical Case (Exp.3).



(NB Error bars indicate +/- 2 standard errors)

If participants were responding logically to the problem, it should be expected that the TA and FC cards would elicit the highest positive rating scores as these cards (irrespective of whether they match or not) are the correct falsifying choices and are therefore the most likely to falsify the rule. This is clearly not the case, in fact it is clear that there is a preference for matching rather than mismatching cards when it comes to their perception in terms of being able to falsify the rule, irrespective of the logical requirement of the task. Unusually, in both conditions the matching TC cards are seen as less likely than the matching FC cards to be able to falsify the rule, even though it can be seen from Figure 4.4 that participants are far more inclined to select the matching TC cards. There would appear to be a discrepancy between the proportion of cards selected and the average ratings that are given.

The combined ratings for both matching and mismatching cards were calculated for each participant in each condition. An average rating score close to zero would indicate a generally reduced perception of relevance attributed to the cards. This could transpire as either a persistent lack of confidence (close to zero ratings for each card) or inconsistent likelihood ratings that swing from the positive (likely) to the negative (unlikely) and therefore cancel each other out. The experimental prediction was that by using semantically related alternative concepts, reasoners' perception of relevance for matching cards would diminish as the alternative card choices might also be seen as relevant. Thus, no specific card would necessarily appear as more relevant than any other. A one-sample t-test was calculated on the averaged rating scores for both matching and mismatching cards to see if they were significantly different from zero (The one-tailed test predicted that ratings would be greater than zero (*likely*) for matching cards, but less than zero (*unlikely*) for mismatching cards). Table 4.10 shows the t-test values and their level of significance.

Table 4.10. T-test Values of Likelihood Ratings for both Matching and Mismatching Cards.

	Matching Card Ratings	Mismatching Card Ratings
Similar[a]	0.94	-0.77
Dissimilar[a]	1.90*	-2.76*
t' - *between*[b]	-0.79 (*df*=60)	1.46 (*df*=60)

*denotes significance

[a] *one-tailed* test for significance (P<.05) from zero

[b] *two-tailed* test for significance (P<.05) between conditions

It can be seen there was a significantly high score of positive (likely) ratings for the matching cards in the *dissimilar* condition but not in the *similar* condition. Likewise, there was a significant amount of negative (unlikely) ratings for the mismatching cards in the *dissimilar* condition but not in the *similar* condition. The difference between groups for mismatching card ratings achieved a near significant value of $t = 1.46$, $df = 60$, $p = 0.074$. It therefore appears that by using semantically similar mismatching cards, participants' perception of each card's relevance has diminished. It would seem that on average, they no longer appear to see either the matching or mismatching cards as being significantly likely or unlikely to falsify the rule. It could be argued that perceptions of relevance in the *similar* condition have been spread across all four cards whereby no one card in particular stands out as being more relevant than the other.

Discussion

Card Selections

The selection patterns for all four cards showed typical quantities of selected cards on a selection task using implicit negations. There were predominantly more matching and mismatching cards selected for TA and TC cases which is reflective of a confirmation rather than falsification strategy. It is even more apparent from the pattern of selections that matching cards are a preferred choice irrespective of the logical case that is represented by the card. These typical matching bias findings replicate the results from Experiments 1 & 2 using materials that do not match at a lexical level. Matching bias occurs even with semantically related items.

There was no significant difference in the rates of matching card selections in both conditions using similar and dissimilar materials. The actual selection frequencies follow the predicted pattern though, in that for all cases apart from TA cards, there were less matching cards selected in the *similar* condition. Likewise, there were more mismatching cards selected for FA, TC and FC cards in the *similar* condition. The AMI and CMI t-values followed this trend in that they showed more matching bias taking place in the *dissimilar* group. Although the mental activation model predicted this trend, it was suggested that card selections alone may be an insensitive measure of the perceived relevance of the cards. Selected cards offer no direct information about relevance judgements of non-selected cards.

It was for this reason that the rating scale was carried out. The rating scale results showed clear differences in the perceptions of relevance of matching and mismatching cards for the similar and dissimilar groups. It was clear that there was a much larger variance in average rating scores in the dissimilar group. It would seem that the cards in the dissimilar group were being seen as far more likely/unlikely to falsify the premise. In the *similar* condition, relevance perceptions of the cards is reduced. Also in the *similar* group, matching cards are seen as less likely to falsify the rule, whereas mismatching cards are seen as more likely (less unlikely) to falsify the rule. There were no significant *likelihood* scores for both the matching and mismatching cards in the *similar* group. It would seem that the relevance of one card being seen as higher than another has been reduced by using semantically similar concepts throughout the cards.

These results offer a new insight into the matching bias phenomenon and provide strong support for the mental activation model. No other existing theory can account for these effects of manipulating the similarity of alternative card concepts. The current explanations of reasoning biases would find it difficult to account for, given the premise clause 'If there is a fruit on the left side...', why the antecedent cards of 'apple' and 'potato' would reduce perceptions of relevance than being given the cards 'apple' and 'glass'. The FA option in both cases is clearly 'not a fruit'. The only possible explanation is that in the first instance 'potato' is semantically related to the superordinate category of 'vegetable' (which would have been the consequent premise clause), whereas 'glass' is not. The relationship of all the cards to *both* the premise items would appear to affect their perceptions of relevance.

It would seem that when all the cards are semantically related to either the antecedent or consequent premise item, the mental activation of the card concepts overlaps the activation in memory of one of the premise items. This pattern of overlapping activation of all the cards makes them appear almost as equally relevant, in line with the predicted mental activation model. Therefore no one card in particular appears as more or less relevant than any other card. Perceptions of whether a card would be likely or unlikely to falsify the rule therefore diminish.

Card Selections and Likelihood Ratings

It was stated at the outset that actual card selections might not be a sensitive enough measure of relevance judgements. The present results would tend to support this contention. Whilst the proportions of card selections appear quite similar between conditions, the rating scores show clear differences. What are also of interest are the rating scores compared to the actual cards selected. Whilst the likelihood rating scores for the TA cards reflect the fact they are predominately selected, whether they match or not, the scores for the other cards do not.

The matching TC cards in both conditions elicit the second lowest amount of 'likely to falsify' ratings of all matching cards, yet they are selected almost as often as the TA cards in both conditions. The mismatching TC cards are the second most popular card selection of all mismatching cards, yet the overall likelihood rating for them lies in the 'unlikely to falsify' group. The discrepancy between selections and likelihood ratings may stem from a confirming rather than falsifying strategy of card selections. Even though participants in both groups rate a card such as the matching FC as likely to falsify the rule, they often do not select it. The FC card is the correct card to choose to falsify the rule. The proportion of TA and TC cards selected suggest a confirmation rather than falsifying strategy. It would seem therefore that the rating scale is not only more sensitive to perceptions of relevance, but may also show a more accurate picture of participants' judgements concerning the ability of the cards to falsify the rule.

Summary

The three selection task experiments have raised two main issues concerning matching bias and reasoners' perceptions of card relevance. It appears that items in a problem do not have to be lexically identical in order to create the matching bias effect. As long as terms on the cards represent those mentioned in the premise, then this would be sufficient for matching bias to occur. Secondly, Experiment 3 has suggested that card selections may not be a true indicator of relevance judgements. There was found to be a poor correlation between the amount of cards selected and the ratings that participants gave to those cards. If the rating scale is a true measure of relevance, then using semantically similar mismatching cards would seem to interfere with relevance attributions. These were novel findings and have in part offered strong support for the proposed mental activation model. The focus in the next two experimental chapters will now turn to the conditional inference task.

To further clarify the present situation with regard to between condition differences in either relevance perception or probability differences of the materials, the conditional inference task may offer an even more precise insight. The conditional inference task controls for one set of outcomes at a time, which can either be accepted or rejected by participants, whereas the selection task contains several outcomes whose relevance can be compared within the actual task. The present studies have aided the defining of matching and mismatching in the selection task and have indicated that there may be disparity between relevance judgements and card selections. Now that it has been shown that matching bias is elicited with semantically related materials, the following chapter will explore the effects of semantic materials in the sentence verification and conditional inference tasks and test further predictions of the mental activation model.

Chapter 5

Experiments 4 & 5 – Conditional Inference Task

The experiments outlined in this chapter set out to address issues relating to both the sentence verification task and the conditional inference task. The sentence verification task used by Wason (1959, 1961) is of a very similar format to the categorisation reaction time experiments of Landauer and Freedman (1968) and Collins and Quillian (1970) – outlined in Chapter 3, *Category Size & Memory.* Essentially in the sentence verification task, participants are shown a sentence of the type '2 is an even number' and are required to give a 'true' or 'false' response. The conditional inference task on the other hand involves giving participants three sentences: a major premise, minor premise and conclusion e.g.

If there is an animal on the left then there is a mineral on the right. ***(Major premise)***

There is a reptile on the left. ***(Minor premise)***

Therefore there is a mineral on the right. ***(Conclusion)***

Participants must then decide if the given conclusion 'follows' or 'does not follow' from the information in the major and minor premises.

Primarily, it is of interest to see if effects of category size manipulation that have been found in previous memory research (outlined in Chapter 3) using the sentence verification type tasks can be replicated using the materials that will also be used for the conditional inference task. Secondly, it would be interesting to see if any effects on the sentence verification task are actually comparable to the conditional inference task. Finally, the study sets out to discover whether the category size manipulation can affect matching bias responses normally found using the conditional inference task.

Matching Bias

In order to study the possible underlying mechanisms that may be responsible for various biased responses in the conditional inference task, the structure of the task can be broken down in several ways which unfortunately has made the area rather jargon-laden. A more complete technical outline was given at the beginning of Chapter 3 and will be recapped briefly here. Using the above conditional inference example, the item that is named in the minor premise is referred to as the *categorical premise*, in this case 'reptile'. The part of the major premise that the categorical premise refers to is called the *referred clause* which in the above example is 'animal'. The item mentioned in the conclusion refers to the part of the major premise known as the *inferential clause* ('mineral'). It is the way the items in the conditional inference task refer to each other that is of particular interest with respect to matching bias. They can be conceptualised as forming two pairings: *referred clause/categorical premise* and *conclusion/inferential clause*.

Matching bias is deemed to occur when there are found to be more endorsements of conditional inferences whose categorical premise and referred clause contain items that match. As was seen in Chapter 4, matching can be defined when a concept semantically *represents* another concept but with explicit negation being ignored i.e. 'oak' represents 'tree', but 'adder' is not seen to represent 'not a tree'. There are several ways at looking at the matching bias effect with conditional inference evaluation tasks. Schroyens, Verschueren, Schaeken and d'Ydewalle (2000) and Evans and Handley (1999) judged matching bias as the difference in the amount of endorsements between inferences with matching categorical premise/referred clause pairings and inferences with mismatching parings. For instance, a matching conditional inference would be of the form given in the above example whereby the categorical premise 'reptile' matches the referred clause concept of 'animal'. A *mismatching* conditional inference which maintains the logical structure of the problem would involve changing the polarity of the referred clause such that;

If there is NOT an animal on the left then there is a mineral on the right.

There is a tree on the left.

Therefore there is a mineral on the right.

It can now be seen that the concept 'tree' mismatches the concept of 'animal'. The use of these implicit negations will be used throughout the study. By implicit, it is meant that the minor premise *implicitly* represents the negation 'not an animal' in the context of the problem e.g. 'tree'. (An explicitly negated minor premise clause would have stated verbatim 'not an animal').

Using only implicit negations in the *negations paradigm* (systematic placement of negations within the major premise), 16 conditional inference problems can be created. These are seen in the reproduced Table 3.1 from Chapter 3.

<u>Table 3.1 (from Chapter 3).</u> The 16 Negations Paradigm Conditional Inferences (Implicit Negation)[2]

	Referred Clause	Inferential Clause	Categorical Premise	Conclusion	Inference Type	Inferential Negation	Referential Negation	Double Negation	Matching
Explicit	A	2	A	2	MP1				•
Affirmation	A	NOT 2	A	NOT 2	MP2	•			•
	2	A	2	A	AC1				•
	2	NOT A	2	NOT A	AC3	•			•
Explicit	NOT A	2	A	NOT 2	DA3		•		•
Denial	NOT A	NOT 2	A	2	DA4	•	•	•	•
	NOT 2	A	2	NOT A	MT2		•		•
	NOT 2	NOT A	2	A	MT4	•	•	•	•
Implicit	NOT A	2	B	2	MP3		•		
Affirmation	NOT A	NOT 2	B	NOT 2	MP4	•	•		
	NOT 2	A	3	A	AC2		•		
	NOT 2	NOT A	3	NOT A	AC4	•	•		
Implicit	A	2	B	NOT 2	DA1				
Denial	A	NOT 2	B	2	DA2	•		•	
	2	A	3	NOT A	MT1				
	2	NOT A	3	A	MT3	•		•	

It can be seen from the table that only half of the inferences contain negated minor premise clauses. Only these can be classed as *implicit*. The other eight inferences contain affirmative minor premises, which are by their very nature explicit representations of the major premise clause. Therefore, even though implicit negations are used throughout the negations paradigm, half of the problems can be classed as explicit and the other half implicit. There is therefore a direct relationship between explicit and affirmative premises, and implicit and negative minor premises. It was seen in Chapter 3 that various authors have interposed both types of labelling. For instance, Schroyens et al classified the inferences by way of two dimensions (as can be seen below in Table 3.1): *referencing* (explicit/implicit) and *inference* (affirmation/denial) whereas Evans and Handley (1999)

[2] The column labelled 'inference type' contains the inference i.e. MP (*modus ponens*) but is appended by a number. This number represents which of the four negations-paradigm premises the inference is made up from. E.g. If the letter is an A then the number is a 2 (1), If the letter is an A then the number is not a 2 (2), If the letter is not an A then the number is a 2 (3), If the letter is not an A then the number is not a 2 (4).

have discussed minor *premise polarity* (affirmative/negative). Minor premise polarity and referencing are essentially the same thing when using only implicit negations.

Matching bias is observed in the difference between explicit and implicit inferences whilst controlling for the logical requirement of the task (affirming or denying). So that matching bias is found when there are more endorsements of explicit than implicit *affirmation* inferences, and more endorsements of explicit than implicit *denial* inferences. Evans and Handley (1999) used an 'Affirmative Premise Index' (API) in order to see if reasoners on inference tasks were more inclined to endorse a conclusion arising from an affirmative rather than a negative minor premise. When used with implicit negations, this index is a measure of *matching bias* between the categorical premise and referred clause. Looking at Table 3.1, it can be seen that when the categorical premise is affirmative it also matches, but when it is negative it mismatches. The API is calculated by counting the number of endorsements of inferences with affirmative minor premises (matching) and subtracting the number of endorsements for inferences which have negative minor premises (mismatching). This index subsequently collapses affirmation and denial type problems.

It would be useful at this point to introduce some shorthand notation that can be used for discussing matching and mismatching 'pairs'. It was pointed out at the end of Chapter 3 (*Memory Activation Model of Reasoning Biases*) that the pairs of interest are the *categorical premise/referred clause* pairing and the *conclusion/inferential clause* pairing. Rather than write the categorical premise clause followed by the referred clause, both clauses will be written separated by an asterix. So that, '2*not2' represents the categorical premise/referred clause pairing from the inference, 'If there is an A on the left there is <u>not a 2</u> on the right. There is a <u>2</u> on the right.' This notation can also be used for the conclusion/inferred clause pairings.

If the matching bias effect can be seen occurring between the categorical premise and referred clause, then it may not be too unreasonable to assume that matching bias could also occur between the conclusion and inferential clause pairing. Essentially, all these pairings match because explicit negation is always used for the conclusion (the problems wouldn't make sense if implicit negation were used). From Table 3.1 it can be seen that there are three types of pairings of conclusion/referential clause that could take place in all 16 negations paradigm inferences: e.g. '2 * 2', '2 * not 2' and 'not 2 * not 2' (and likewise for the antecedents). Considering matching bias effects for these clauses is not too dissimilar to the 'matching$_2$' effect proposed by Oaksford (2002) to explain explicit negation effects in the selection and truth table tasks. He proposed that this additional form of matching$_2$ bias might be a strategy taken by reasoners, for instance in the selection task in which they may match the whole clause on the card to the clause mentioned in the premise, including negations (see Chapter 1, *Implicit Negation*).

These three pairings (no negations, one negation and two negations) were incorporated into the mental activation model outlined in Chapter 3. Previous researchers have looked at possible conclusion/referential clause pairing effects but only generally in terms of the polarity of the conclusion. For instance, Evans and Handley (1999) used a 'Negative Conclusion Index' (NCI). This is calculated by counting the number of endorsements of inferences with negated conclusions and subtracting the number of endorsements of inferences with affirmative conclusions.

Looking at the pairings of *conclusion* and *inferential clause*, it can be seen that affirmation problems make use of two types of pairing; '2*2' and 'not 2*not2'. Whereas the denial problems only use one type of pairing; '2*not2' (or 'not2*2'). It could be that analyses using the NCI calculations may in some way detect a form of matching bias (e.g. '2*2', 'not2*not2'). The conclusion *clause* in affirmation problems always directly matches the inferential clause, whereas in denial problems the conclusion partially matches (i.e. '2*not2'). The mental activation model describes how the ''2*2' and 'not2*not2' pairings completely overlap in activation and might therefore appear equally as relevant (therefore a non-significant NCI). The '2*not2' pairing only partially overlaps in the activation of the concepts and might therefore be seen as less relevant. Although this does not directly explain the preference found by Evans and Handley for inferences with negative conclusions in denial problems but not affirmation problems, it may offer part of the explanation in terms of a bias caused by relevance perception of the problem materials.

The API scores effectively differentiate between explicit and implicit type inferences and the NCI scores generally only show differences of conclusion polarity on denial problems. The NCI on denial inferences was said by Evans and Handley (1999) to be indicative of a 'double negation' effect whereby reasoners have difficulty with four of the eight denial inferences that consist of affirmative premises (see Table 3.1). On these problems, reasoners may be led to deny the inferential clause. The inferential clause in all the denial problems with affirmative conclusions are in fact negative. Therefore, the double-negation effect occurs when trying to work out what a 'not not a 2' or 'not not an A' represents.

API scores will be used in the present study as a measure of matching bias between the categorical premise and referred clause, and the NCI scores will be included for comparison purposes. It will be useful for the present study to analyse the amount of endorsements given to the inferences using the Schroyens et al (2001) divisions of explicit/implicit and affirmation/denial type inferences. If matching is taking place between the pairings 'categorical premise/referred clause' and 'conclusion/inferential clause' then matching bias on the former pairing will show as more endorsements for *explicit* problems and matching for the latter pairing will show as more endorsements for *affirmation* problems. (On affirmation problems the conclusion/inferential clause concepts exactly match e.g. 'A*A','2*2', 'not A*not A' and 'not 2*not2', whereas on denial problems they only partially match e.g. 'A*notA', '2*not2' and vice-versa).

The mental activation model outlined in Chapter 3 for the conditional inference task highlights the prediction that by perceived relevance alone, there will be the following proportions of inference endorsements: Exp.Aff > Imp.Aff. > Exp.Den. > Imp.Den. The mental activation model takes into account matching between categorical premise/referred clause *and* matching between conclusion/inferential clause. Analyses can be broken down by the divisions described above (implicit/explicit affirmation/denial) for both matching bias with the categorical premise/referred clause and possible matching with the conclusion/inferential clause.

Experiment 4

Participants

A total of 64 undergraduate students from the University of Plymouth participated, either as paid volunteers or as part of their course requirement. Participants were tested in groups of up to six at a time on separate computers and were randomly, and equally, allocated to either of the two conditions (condition 1 n=32; condition 2 n=32).

Design and Procedure

The experiment manipulated three levels of concept size *pairings*. These were configured for all inferences using the three concept sizes: large, medium and small from *nested triple* categories. For example, a nested triple 'animal/reptile/adder' assures that the concepts decrease in size because the smaller concepts are actual inclusions of the larger concepts. Thus it is not possible that 'reptile' is larger than the concept 'animal' because 'animal' includes 'reptile'. The pairings were thus: *(medium * large)*, *(small * large)* and *(small ** *medium).* These *pairings* will from now on be referred to as *m-l*, *s-l* and *s-m* respectively, because of the combined size of the two concepts that form the pair.

There were two parts to the experiment: a sentence verification task and a conditional inference evaluation task. Six sets of nested triple categories were used: *animal/reptile/adder, plant/tree/oak, food/cake/eclair, mineral/gem/diamond, activity/sport/tennis and building/gallery/the Tate* (the interaction of materials and inferences can be seen in Appendix D) from which matching and mismatching paired concepts are drawn. A 'pair' describes the two items used for the categorical premise and referred clause. (As was pointed out earlier, the conclusion/referential clause items can not be formed with category items, as this would interfere with the logic of the task. They therefore have to display the same verbatim concept – with or without negation). The sentence verification task in this study also utilises the same two items used for the conditional inference task pairs. To make a *matching* pair, concepts from the same nested triple were used, and to make a '*mismatching pair*' concepts from different nested triples were used. Examples of the *matching* sentence verification and corresponding conditional inference problems are shown in Table 5.1. The table also shows how they form the experimental conditions.

Table 5.1. Examples of the Matching Materials Used for Each Condition in Experiment 4.

	Category Pairing 1	Category Pairing 2
Sentence Verification Task		
m-l/s-l Group	a *reptile* is an *animal* (m-l)	an *adder* is an *animal* (s-l)
s-m/s-l Group	an *adder* is a *reptile* (s-m)	an *adder* is an *animal* (s-l)
Conditional Inference Task		
m-l/s-l Group	- If there is an *animal* on the left side then there is a *mineral* on the right side. - There is a *reptile* on the left side. - Therefore there is a *mineral* on the right side.	- If there is an *animal* on the left side then there is a *mineral* on the right side. - There is an *adder* on the left side. - Therefore there is a *mineral* on the right side.
s-m/s-l Group	- If there is a *reptile* on the left side then there is a *mineral* on the right side. - There is a *adder* on the left side. - Therefore there is a *mineral* on the right side.	- If there is an *animal* on the left side then there is a *mineral* on the right side. - There is an *adder* on the left side. - Therefore there is a *mineral* on the right side.

Concept sizes were manipulated both between and within conditions. Such that, in the *m-l/s-l* group, participants were given 16 cases (for each of the sentence verification task and conditional inference task) in which the category pairings were *m-l* and 16 cases where the category pairings were *s-l*. In the *s-m/s-l* group, participants were given 16 cases where the category pairings were *s-m* and 16 cases where the category pairings were *s-l*. Thus each participant was given 32 sentence verification tasks and 32 conditional inference evaluation tasks. The mixed design was used in order to give the analysis more power and at the same time avoiding the fatigue for participants of doing all three conditions (96 tasks) in one experiment. Here the total number of tasks (sentence verification and conditional inference) is reduced to 64 per participant.

The same category pairings were used in both parts of the experiment. The first part, the sentence verification task, required participants to decide if the displayed sentence *follows* (true) or *does not follow* (false).

Four sentence types were used e.g.:

> *2 is an even number (TA)*
>
> *2 is an odd number (FA)*
>
> *2 is not an even number (FN)*
>
> *2 is not an odd number (TN)*

Thirty-two sentences were displayed (eight of each sentence type), 64 in total.

The second part of the experiment, the conditional inference evaluation task, used the same concept pairings as in the sentence verification task for the categorical premise/referred clause. Inferential clause/conclusion pairings were not mirrored in the sentence verification task, because the structure of the inference precludes the use of category pairs for these concepts i.e. the conclusion item is the inferential clause item, with or without a negation (see Table 3.1). In the inference task, participants were again required to evaluate whether the given conclusion *follows* or *does not follow*.

Instructions

The instructions for the sentence verification task were:

> *The first part of this experiment involves making a decision about sentences. There are 32 sentences in total and you will only see one at time. All that is required is to judge whether the single sentence that appears on the screen is true 'FOLLOWS' or false 'DOES NOT FOLLOW'.*
>
> *There are only two possible responses that can be made. These are: 'FOLLOWS' and: 'DOES NOT FOLLOW'. These responses are activated by the two coloured keys that are marked on the keyboard. You will see that the key on the right is marked in red ('FOLLOWS'), the key on the left is marked in blue ('DOES NOT FOLLOW').*
>
> *An example of a sentence would be: "a cat is not a mineral". The answer in this case being 'FOLLOWS'. Do not be surprised if the majority of the sentences you see are as bizarre as this one - all you have to do is answer 'FOLLOWS' or 'DOES NOT FOLLOW' as quickly as possible.*
>
> *Press 'C' to continue.*
>
> *Please remember to respond as quickly as possible, but at the same time, try to avoid making mistakes.*
> *If you have any questions, then please ask BEFORE you commence the experiment.*
>
> *When you are ready, press 'C' to start the experiment.*

The instructions for the conditional inference evaluation task were:

> *This final part of the experiment involves a series of 32 problems. Each problem consists of a rule followed by two statements. Your task is to decide whether the two statements that follow the rule actually obey the rule.*
>
> *Again, there are two possible responses that can be made. These are: 'FOLLOWS' and 'DOES NOT FOLLOW'.*
> *These responses are activated by the two coloured keys that are marked on the keyboard. You will see that the key on the right is marked in red ('FOLLOWS'), the key on the left is marked in blue -('DOES NOT FOLLOW').*
>
> *All of the 32 rules concern items that appear on fictitious sets of cards. You are to imagine that the cards have two pictures printed on them - one picture on the left side of the card and one picture on the right side of the card. There are ONLY two pictures on each card - one on the left and one on the right.*
> *The two statements that follow the rule concern possible combinations of pictures. Your task is to judge whether the two statements that follow the rule actually obey the rule ('FOLLOWS') or do not obey the rule ('DOES NOT FOLLOW').*
>
> *Press 'C' to continue.*
>
> *Please remember to respond as quickly as possible, but at the same time, try to avoid making mistakes.*
>
> *If you have any questions, then please ask BEFORE you commence this part of the experiment.*
>
> *When you are ready, press 'C' to start the experiment.*

The response latencies and actual response (*follows/does not follow*) were recorded in both parts of the experiment. All tasks were displayed on a computer screen where participants were required to give a response by pressing one of two coloured keys on the keyboard. The problems were randomised by the computer program within each part of the experiment. Each participant received tasks that were from one of 16 blocks. The materials and problem types (16 negations paradigm inferences) were randomly associated to make up each block of tasks. This controlled for any anomalous interaction between materials and type of task (i.e. TA, FA, TN, and FN in the sentence verification task, and negation paradigm/logical case in the inference task).

As the pairs used for the sentence verification task mirrored those used in the conditional inference task, conditional inference task terminology will be used here for the analysis at it avoids confusion when comparing the materials between both parts of the experiment. The pairs of concepts used for the four sentence verification types (TA, FN, TN and FA) reflect the concept pairs used for the four inference types (explicit affirmation, explicit denial, implicit affirmation, and implicit denial) in the following way: explicit affirmation=TA, explicit denial=FN, implicit affirmation=TN, and implicit denial=FA. For instance, using the following *explicit denial* inference, 'If there is *NOT an animal* on the left side then there is a mineral on the right side. There is a *reptile* on the left side. Therefore there is NOT a mineral on the right side', it can be seen that the categorical premise/referred clause concepts are 'reptile' and 'not an animal'. When these concepts are included in a sentence verification task format it produces the following, 'a *reptile* is *not an animal*'. This is now an example of a false-negative (FN) sentence requiring the response of 'false'.

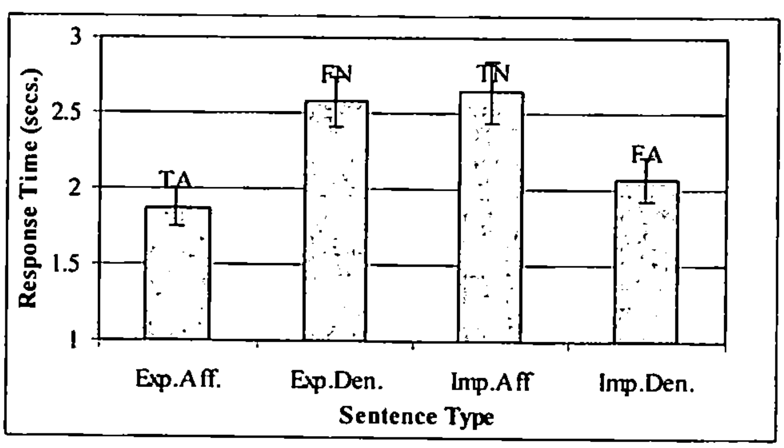
Results of the Sentence Verification Task

Response Latencies

A 2x2x2x2 mixed ANOVA was calculated on the response times, where the between group factor was *concept size* (m-l/s-l & s-m/s-l concept pairings) and the within group factors were *referencing* (explicit/implicit), *inference* (affirmation/denial) and *category pairing* (m-l/s-m& s-l). Response times for the four sentence types were combined (explicit affirmation/*TA*, explicit denial/*FN*, implicit affirmation/*TN*, and implicit denial/*FA*). Outliers were omitted from the data on the criteria of being two standard deviations below, and three standard deviations above, the overall mean. This meant that combined response times outside of the range of 6-21 seconds were excluded (one case from condition 1 and two cases from condition 2).

There was found to be a main effect of *reference* ($F_{1,59} = 5.21$, $MSE = 6.30E6$, $p<0.05$). The means (explicit 8.89 and implicit 9.41 sec) show that explicit sentences (TA & FN) took significantly less time to respond to than implicit sentences (TN & FA). Explicit referencing refers to when both of the items mentioned in the sentence match (i.e. oak & plant) as opposed to mismatch (i.e. oak & building). The main effect of reference supports the matching bias prediction in that it would seem participants are quicker to respond to sentences in which both items match, irrespective of the requirement of the task. Such that, sentences of the form '*an oak is a plant*' (TA) are quicker to respond to than '*an oak is a building*' (FA), and sentences of the form '*an oak is not a plant*' (FN) are quicker to respond to than '*an oak is not a building*' (TN). There was also a significant interaction between *reference* and *inference* ($F_{1,59} = 151.32$, $MSE = 5.19E6$, $p<0.01$) which can be seen in Figure 5.1. (Inference in this case refers to the truth or falsity of the sentence). These results reflect previous findings using the sentence verification task in that the response latencies follow the pattern; explicit affirmation<implicit denial<explicit denial<implicit affirmation.

Figure 5-1. Interaction of Mean Response Times for Reference Type and Inference (Exp.4).

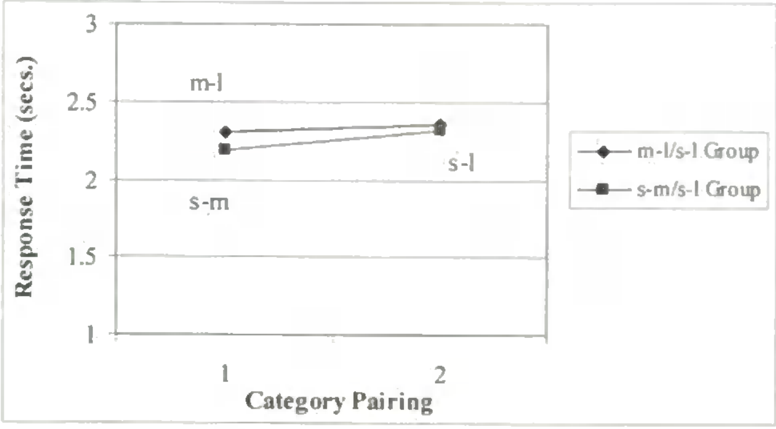


(NB Error bars indicate +/- 2 standard errors)

Due to the mixed design, it was expected that any influence of concept size would be apparent in the ANOVA in the form of an *interaction* between *concept size* and *category pairing*. This somewhat complex design needs further explanation. In each group the participants were given two sets of sentences made up of different concept sizes. Participants in each condition were given sentences made up of *s-l* concept pairs which was labelled *category pairing 2* for the analysis. *Category pairing 1* was different in both groups, with the *m-l/s-l* group, category pairing 1, being constructed from the *m-l* concept pair. In the *s-m/s-l* group, category pairing 1 was constructed from the *s-m* concept pair.

The interaction between *category pairing* and *concept size* was not significant. Figure 5.2 only just shows that the pattern of responses followed the prediction for the s-m/s-l group, in that the *s-m* concept pairing produced faster response times than sentences in which *s-l* concept pairs were used. A follow-up LSD analysis showed that there was a significant difference between *s-m* (s-m/s-l group) and *s-l* (m-l/s-l group) with $p < .05$. It would seem that concept size produced a very weak overall effect in the sentence verification task and did not reproduce the stronger findings from previous categorisation reaction time studies. Previous research found a stronger category size effect for denial sentences (FA) so separate t-tests were carried out on just the 'implicit denial' (FA) sentences. No significant differences were found between conditions and so the concept size manipulation used in this study was not seen to replicate previous findings in the literature.

Figure 5-2. Interaction of Mean Response Times for Condition and Category Level (Exp.4).



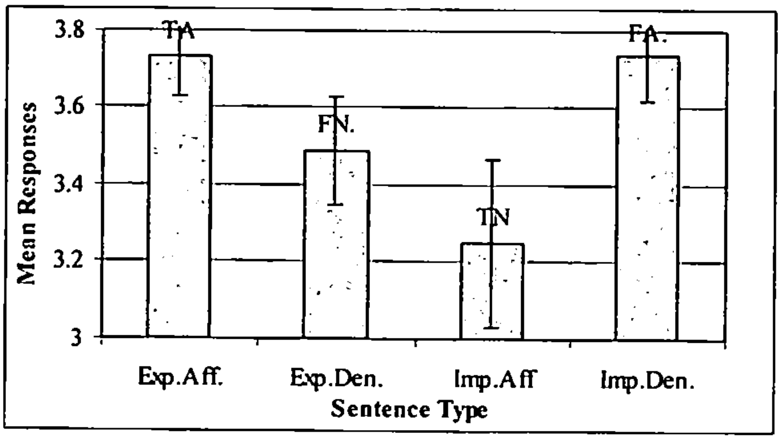
Responses

The responses given to the sentence verification task will be looked at in two ways. First, the amount of 'follows' responses (endorsements) can be analysed to reveal if there were any forms of matching bias taking place. Matching bias would be revealed when there are generally more 'follows' responses for sentences with matching items (TA & FN) than for mismatching items (TN & FA). The second form of analysis will focus on the amount of correct responses given to the sentences (i.e. TA & TN = *follows*, FA & FN = *does not follow*).

Using the same ANOVA design as was used for the response latencies, there were found to be no significant effects of *reference* or *inference*, or even a significant interaction between *category pairing* and *concept size*. This was not surprising considering the overall large amount of correct responses that were given to the task. A 'follows' response to implicit denial (FA) and explicit denial (FN) type sentences would be classed as incorrect.

The average mean amount of 'correct' responses for each of the four sentence types was 3.55 out of a possible 4. An ANOVA on the amount of correct responses revealed a significant interaction of *referencing* and *inference* ($F_{1,62}$ = 35.32, *MSE* = 0.489, p<.01) which can be seen in Figure 5.3.

Figure 5-3. The Interaction of Reference and Inference on Correct Responses in the Sentence Verification Task (Exp.4).



(NB Error bars indicate +/- 2 standard errors)

Explicit affirmation (TA) and implicit denial (FA) sentences elicited the largest proportion of correct responses. It will be noted that these two sentence types are the ones that do not contain a negation (i.e. the word 'not'). These results may reflect possible processing difficulties associated with the presence of a negation, whilst the overall larger proportion of correct responses for 'explicit' sentences could be a reflection of matching bias, in that matching items would seem to facilitate correct responding.

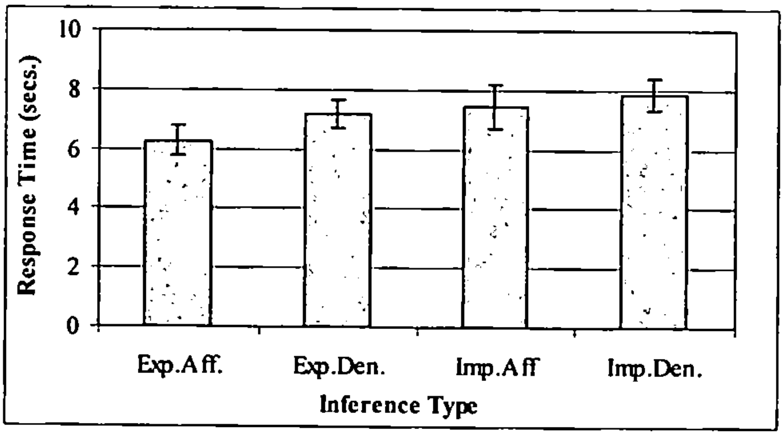
Results of the Conditional Inference Task

Response Latencies

A 2x2x2x2 mixed ANOVA was again calculated on the response times using the same design as before. To gain a more accurate measure of the actual response times taken for the inference task, the combined times for each of the four sentence types (i.e. explicit/implicit affirmation/denial) on the verification task were taken away from the respective times for the conditional inference task. This should control for reading times and general comprehension, leaving a pure latency measure for making the actual inference.

Outliers were again omitted from the data on the criteria of being two standard deviations below, and three standard deviations above, the overall mean. This meant that only combined response time differences (inference minus verification task) in the region of 4 to 73 seconds were included in the analysis, which excluded one case from each group. The analysis revealed two significant main effects of; *referencing* ($F_{1,60}$ = 26.60, *MSE* = 6.791E7, $p < .01$) and *inference* ($F_{1,60}$ = 13.08, *MSE* = 6.841E7, $p < .01$). Looking at Figure 5.4 it can be seen that these effects were mainly caused by the faster mean response times to explicit affirmation problems (mean = 6.25sec).

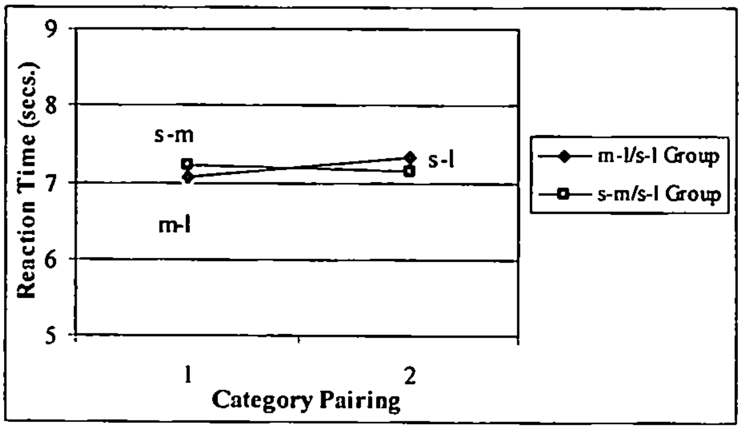
Figure 5-4. Mean Response Times by Referencing and Inference (Exp.4).



(NB Error bars indicate +/- 2 standard errors)

The ANOVA revealed no significant interaction of *concept size* and *category pairing*. It can be seen from Figure 5.5 that there is very little difference in response times between the factors. It would appear that the concept size manipulation, as in the sentence verification task, did not produce a significant effect on inference response latencies.

Figure 5-5. Mean Response Times by Concept Size and Category Pairing (Exp.4).



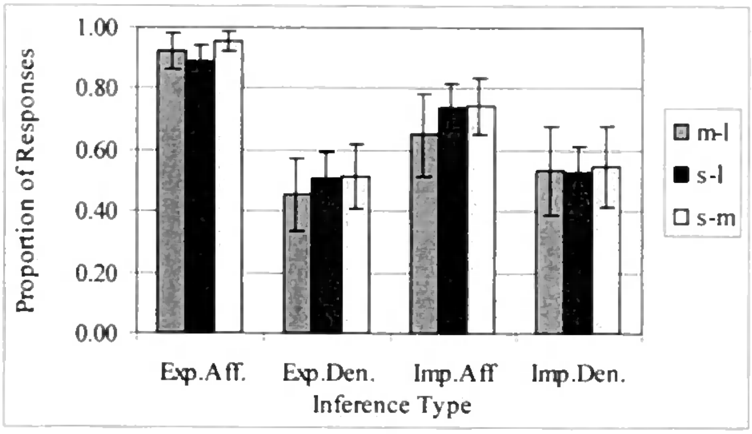
Responses

The 2x2x2x2 mixed ANOVA design was used as before to analyse the responses. The mixed ANOVA revealed a significant main effect of *referencing*, $F_{1,62} = 10.67$, $MSE = 1.231$, $p < .01$. This essentially measures the same responses as the API (Evans & Handley, 1999). The API is a measure of categorical premise/referred clause matching_1, which coincides with minor premise polarity. As there were more endorsements of explicit inferences there is seen to be a significant tendency to endorse more inferences whose categorical premise matches the referred clause.

There was a also a significant main effect of *inference* in that more affirmation inferences were endorsed than denial inferences ($F_{1,62} = 46.55$, $MSE = 4.032$, $p < .01$). In affirmation inferences, the conclusion clause matches_2 the inferential clause i.e. '2*2', and 'not2*not2', whereas the pairings in the denial inferences mismatch ('not2*2' or '2*not2'). This account of matching with the clauses (including negation) fits in with Oaksford's (2000) 'matching_2' account. It is therefore argued that differences in endorsements to affirmation and denial problems are a reflection of a conclusion/inferential clause matching_2 effect.

The proportion of responses to each of the four inference types can be seen in Figure 5.6.

Figure 5-6. The Proportion of Responses for Each Inference Type and Category Pairing. (s-l category pairs averaged across both conditions) (Exp.4).



(NB Error bars indicate +/- 2 standard errors)

Category Pairings

Table 5.2 shows the proportion of endorsements ('follows' responses) by logical case for both category pairings in each concept size group.

Table 5.2. Proportion of 'follows' responses for each logical case.

m-l/s-l Group	Category Pairing '*m-l*'				Category Pairing '*s-l*'			
	MP	DA	AC	MT	MP	DA	AC	MT
If P then Q	94	59	91	63	88	59	91	69
If P then not Q	88	47	59	66	88	47	69	56
If not P then Q	72	41	97	44	72	56	88	53
If not P then not Q	69	41	59	34	69	50	75	38

s-m/s-l Group	Category Pairing '*s-m*'				Category Pairing '*s-l*			
	MP	DA	AC	MT	MP	DA	AC	MT
If P then Q	97	59	91	59	91	59	91	41
If P then not Q	97	63	81	59	94	53	75	63
If not P then Q	69	50	97	38	84	56	84	38
If not P then not Q	72	44	75	53	78	28	69	56

NB Matching cases underlined

An overall effect of the three category pairings would be apparent in a significant interaction between *concept size* and *category pairing*. This did not reach significance ($F_{1,62} = 2.5$, $MSE = 0.612$, $p = .12$), although an LSD post hoc comparison revealed a significant difference between the amount of responses for the category pairings *m-l* and *s-m* ($p < .05$). Figure 5.7 shows that the proportions of endorsements were in the predicted direction.

Figure 5-7. Proportion of Follows Responses for Both Category Pairings in Each Condition (Exp.4).

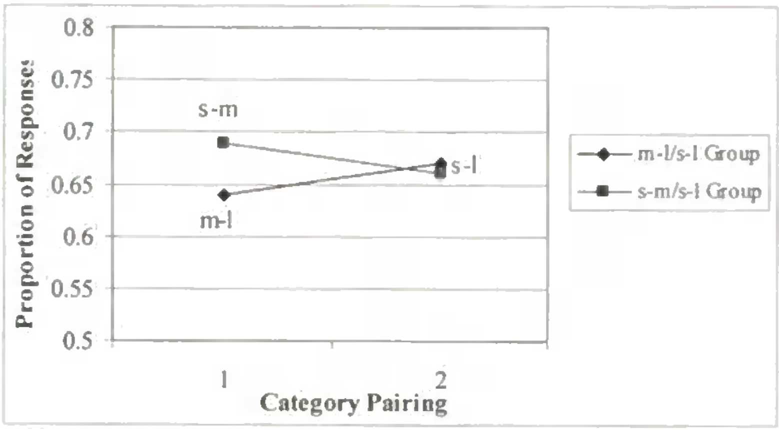


Figure 5.7 shows the mixed design in which *category pairing 1* for the m-l/s-l group were the inferences made up of *m-l* concepts, and for the s-m/s-l group the inferences were made up of *s-m* concepts. For *category pairing 2* for both groups the inferences were made up of *s-l* concepts. It can be seen that the proportions for each of the three category pairing manipulations follow a linear trend in that the *m-l* concept size elicited the least endorsements and *s-m* concept size elicited the most. In both groups, the s-l concept sizes elicited almost equivalent amount of endorsements. Thus, the observed trend (significant difference between *m-l* & *s-m*) for proportions of endorsements and hence the amount of matching bias was *m-l* < *s-l* < *s-m*, which follows the relevance prediction reflected in the memory activation model. (Relevance here being assumed by the amount of inference endorsements).

API

The API is a more sensitive measure of detecting the presence of matching between the categorical premise and referred clause by subtracting the amount of endorsements of inferences with negative minor premises (mismatching) from those with positive minor premises (matching). Table 5.3 shows the one-tailed t-test values used to detect whether each index is significantly different from zero, giving an indication of whether a significant matching bias effect is taking place.

Table 5.3. T-test values for the API Scores for Each Concept Size and Category Pairing (Exp.4).

	Pairing 1	Pairing 2
m-l/s-l Group[a]	1.87 * (*m-l*)	1.34 (*s-l*)
s-m/s-l Group[a]	2.96 * (*s-m*)	1.83 * (*s-l*)
Between Groups[b]	0.13 (df=49)	0.53 (df=60)

* Denotes significance
[a] *one-tailed* test for significance (P<.05) from zero
[b] *two-tailed* test for significance (P<.05) between conditions
Concept size in brackets

It can be seen from Table 5.3 that there was no significant matching bias in the m-l/s-l group with the *s-l* size concepts. There is no clear explanation for this apart from that the materials used in this group were the largest concept sizes overall in the experiment. There may have been carry-over effects of the concept sizes caused by the within-subjects design of the experiment. It was predicted that there would be less matching bias present when these large concept size materials were used and this should have been reflected also with a reduction in matching bias for the *m-l* concepts. It is not clear at present whether the larger sized concepts in the m-l/s-l group were responsible for this finding.

NCI

The Negative Conclusion Indices (NCI) were calculated in line with Evans and Handley (1999) where the amount of endorsements for inferences with affirmative conclusions is subtracted from the amount for negative conclusions. Thus, a positive NCI indicates that reasoners had endorsed more inferences that had negative conclusions. Table 5.4 shows the t-test values and their significance level. The first two rows show the t-test values for a one-sample test to see if the indices were significantly different from zero.

Table 5.4. T-test values for the NCI Scores for Each Concept Size and Category Pairing (Exp.4).

NCI t-values	Pairing 1	Pairing 2
m-l/s-l Group[a]	2.12* (m-l)	2.24* (s-l)
s-m/s-l Group[a]	1.22 (s-m)	1.03 (s-l)
Between group[b]	0.63	0.69
	(df=61)	(df=60)

* Denotes significance
[a] one-tailed test for significance (P<.05) from zero
[b] two-tailed test for significance (P<.05) between conditions
Concept size in brackets

In the *m-l/s-l* group the NCI was significantly different from zero for both category pairings, but were not significant in the *s-m/s-l* group. It would appear that concept size had an effect on both the NCI and API values. The API scores, which are a measure of matching bias between the categorical premise and referred clause, showed a weaker matching effect (no significant effect for the *s-l* concept size) in the m-l/s-l group. The NCI scores show a significant negative conclusion bias for the m-l/s-l group but not for the s-m/s-l group. It almost appears that if there is significant matching (API) taking place then there will not be significant negative conclusion bias. There was no significant difference though between groups.

A 2x2x2x2 ANOVA was calculated on the NCI values, again where the between factor was *concept size* and the within factors were *category pairing*, *reference* and *inference*. The results showed a significant main effect of *inference* ($F_{(1,62)} = 12.93$, $MSE = 0.658$, $p < 0.001$), where the mean NCI for affirmation inferences (mean = -0.02) was much lower than the mean NCI for denial inferences (mean = 0.24). This effect is predicted in line with the double negation effect and mirrors the findings of Evans and Handley (1999). The double negation effect is seen only on denial inferences, which have affirmative conclusions, where reasoners are less likely to endorse these inferences causing the NCI value to be greater. There was no other significant main effect or interaction with any other factors.

Discussion

Sentence Verification Task - Matching Effects

The sentence verification task showed a clear effect of matching bias in that the 'matching' pairs in the sentences were responded to faster than the 'mismatching' pairs. Previous sentence verification task results (as outlined by Evans, 1982) showed a typical pattern of response times of the order TA<FA<FN<TN which was also found here. Separating the four sentence types by the presence or absence of negations, matching can be seen when there are faster TA than FA responses (i.e. 'an oak is a plant' is responded to quicker than 'an oak is a building'). In the presence of a negation, it is seen that there are faster responses to FN sentences than TN (i.e. 'an oak is not a plant' produces faster response times than 'an oak is not a building').

Logical Responses & Latencies

The proportion of correct responses in the sentence verification task mirrored the response latencies, in that the amount of correct responses was inversely proportional to the time taken to give a response. It would appear that matching pairs in the sentences facilitated both the logical response and the time taken to make that response. The response latencies in this case can therefore be considered to be a reflection of the difficulty of the task. The proposed mental activation model predicts these findings in terms of the activation overlap that occurs using materials from the same semantic domain. When materials do not match, there is no mental activation overlap and therefore activation levels do not facilitate the perception of the problem as being seen as relevant.

Negation

There is also another underlying effect concerning the presence or absence of negation. Those sentences that did not contain a negation (TA & FA) produced faster response times and more correct responses than the sentences which did contain a negation (FN & TN). The mental activation model only predicts underlying biases at an unconscious representational stage, which is parallel to Evans' (2002) dual process link with Stanovitch's (1999) System1 processes (Chapter 1, *Heuristic-Analytic & Dual Process*). It is apparent that formal, conscious System2 processes are very much at work as is seen by the large proportion of correct responses across all four sentence-types. It could be argued that the sentence verification task is much simpler than tasks such as those involving conditional inferences and therefore underlying biases to not show up to such a large degree, as the correct answer is more apparent. It could be that biases caused by the actual content of a problem become more apparent when the participant is unsure of the correct answer. The findings that sentences are responded to quicker and with more correct answers when there are no negations present may be indicative of negations causing representational anomalies in memory activation.

Category Size

With regard to the concept sizes represented by the items in the sentence verification task, there appeared to be no significant difference in response times or correct responses, using the three concept size pairings. It was predicted that the pairings made up from smaller concepts would produce faster response times than those made up from larger parings, in line with previous findings from the memory literature. Many of the previous experiments that produced these findings used the materials in a within-subjects design. It could be that in the present experiment, the mixed design did not allow for the two extreme category sizes (*m-l* & *s-m*) to be compared on a within-subjects basis, and therefore was not sensitive enough to detect differences in either response latencies or the amount of correct responses. Perhaps an alternative reason could have been the fact that the experiment did not utilise very many nested categories and therefore was more prone to extraneous influences such as familiarity or frequency of usage of the concepts that were used.

Conditional Inference Task - Latencies

The inference task response times did not seem susceptible to the presence or absence of negations in the major premise. The response times for the inference task were analysed by subtracting the response times taken for the corresponding verification task. This in part, took away individual differences in reading and comprehension times to give a 'purer' measure of the time taken to reason about the inference problem. Figure 5.4 shows that implicit denial problems appear to be the most difficult (if response latency is taken as a measure of problem difficulty). Both types of inference in which the categorical premise item matches the referred clause item (explicit affirmation and explicit denial) appear to be easier inferences on which to decide if the conclusion 'follows' or not. The largest effect then appears to be caused by matching/mismatching pairings. So, unlike the findings from the sentence verification task, the presence or absence of an explicit negation (in the major premise) does not seem to have such a large effect on responses. As previously stated, this could happen because matching bias has a greater effect when there is uncertainty as to the correct response that needs to be given. The response times for the inference tasks did not show up any significant differences using the three different concept size materials and this reflects the findings of the response times for the verification task.

Inference Endorsements & Latencies

The proportion of 'follows' responses for the inference task (explicit affirmation>implicit affirmation>implicit denial>explicit denial) are not inversely proportional to the pattern of response times (except for explicit affirmation problems). Whilst a 'follows' response is not necessarily the correct response for the denial problems, it can be seen from Figure 5.6 that these inferences only elicited roughly 50% 'does not follow' responses (50% '*follows*' = 50% '*does not follow*') – an almost random response rate. Unlike the verification task, the response latencies in the inference task may not be a true indicator of problem difficulty. Clearly, reasoners have difficulty with denial inferences, whereas in the verification task, denial sentences (FA & FN) elicited roughly the same proportion of correct responses as their affirmation counterparts (controlling for the presence of negation).

Matching Bias

Matching bias is seen when there are more endorsements ('follows' responses) to inferences in which the categorical premise and referred clause match rather than mismatch, which in this case is when there are more endorsements to explicit rather than implicit problems. Figure 5.6 clearly shows the overall trend predicted by the mental activation model in that there is a greater matching effect seen on *affirmation* problems (a greater difference in the amount of endorsements between explicit affirmation and implicit affirmation inferences). The mental activation model predicts a greatly reduced matching effect with *denial* problems (difference in endorsements between explicit denial and implicit denial inferences) and it can be seen that there are roughly equal proportions of explicit denial and implicit denial responses. In other words, there is very little matching bias seen when comparing responses between the denial problems.

Concept Size

It was predicted that the larger the activation area represented by the concepts (i.e. larger categories), the less likely matching bias will occur. When looking at the total amount of inference endorsements as an indicator of matching bias, there was no significant main effect, but the proportions do follow the predicted direction in that the *s-m* category pairing elicited significantly more endorsements than *m-l* category pairing. This finding offers support for the mental activation model.

Although the account of Oaksford, Chater and Larkin (2000) prescribes differences in the probabilities of the materials (especially the probability of the conclusion in a conditional inference task) as having an effect on whether reasoners are more likely to endorse the inference, this account does not hold with the materials used in the present study. In the present study, the category size manipulation also alters the probability of concepts used in the inference e.g. something is more likely to be a tree than it is to be an oak. But in terms of the context of the problem, it may be just as probable that a tree could be an oak as it is that a plant could be a tree. These were the types of materials used for the referred clause and categorical premise respectively to represent the *s-m* and *m-l* category pairings. Therefore, in terms of contextualising the categorical premise in terms of the referred clause, there should be no real difference between these materials in terms of their probability of occurrence.

Relevance

The present findings might add clarity to Evans' (1998) contention that reasoners are more likely to endorse problems that they see as being relevant. The mental activation model suggests that relevance judgements are directly related to the level of activation that takes place in memory. These differences in relevance judgements occur both between the four different inference types (explicit/implicit affirmation/denial) as is seen by the usual pattern of matching bias, and can also occur by using different sized concepts.

Experiment 5

The concept size effect seemed to be weak in Experiment 4 as is revealed by the sentence verification results. This may be due to the limited amount of nested triple categories used and also carry over effects caused by the within-subjects design. The same nested triples were used for both the 'within' conditions and it could be that carry over effects occurred because participants were given all three concepts from a nested triple. That is, participants in the *s-m/s-l* group were given the pairings of *tree* and *oak* (*s-m*) and also *oak* and *plant* (*s-l*). In terms of mental activation, there could be trace activations of all three concepts occurring and this might therefore reduce the effect size. They might unconsciously be reading *plant* as representing *tree* or vice-versa. It was therefore necessary to try to replicate these concept size findings in a completely within design, but this time using different nested triples for *each* condition, thereby eliminating the possibility of any carry-over effects caused by mental activation or re-interpretation.

Participants

A total of 40 participants from the University of Plymouth took part as paid volunteers. The experiment took place over the course of two weeks with participants being tested in small groups of up to two at a time.

Design and Procedure

The three category pairings were manipulated as in Experiment 4, but this time they were manipulated in a totally within design. Each participant received a total of 48 problems made up from all three category-pairings (*m-l*, *s-l* and *s-m*), with 16 negations paradigm inferences for each of the three category pairing manipulations. Twenty-four nested category triples were used in total (Appendix E) with eight being used for each category pairing (16 inferences). Three blocks of 48 inferences were created with materials being randomly allocated to both *category pairing* and *inference* (any one of the 16 negations paradigm inferences). This was to control for any possible materials/inference type interaction. Participants were randomly allocated one of the three blocks of inferences. As previously, both response latencies and 'follows/ does not follow' responses were recorded. Examples of matching inferences for the three category pairings can be seen in Table 5.5.

<u>Table 5.5.</u> Examples of Matching Inferences for the Three Category Pairings (Exp.5).

m-l Category Pairing (medium • large)[1]	*s-l* Category Pairing (small •large)[1]	*s-m* Category Pairing (small •medium)[1]
- If there is an <u>*animal*</u> on the left, then there is a *plant* on the right.	- If there is a <u>*living thing*</u> on the left, then there is an *activity* on the right	- If there is a <u>*gem*</u> on the left, then there is a *cake* on the right.
- There is a <u>*reptile*</u> on the left.	- There is a <u>*monkey*</u> on the left.	- There is a <u>*diamond*</u> on the left.
- Therefore there is a *plant* on the right	- Therefore there is an *activity* on the right.	- Therefore there is a *cake* on the right.

[1] Concept size for the referred-clause•categorical premise shown in parentheses

The instructions and procedure for the experiment were identical to Experiment 4 except that the sentence verification part was omitted and there were 48 inference problems instead of 32.

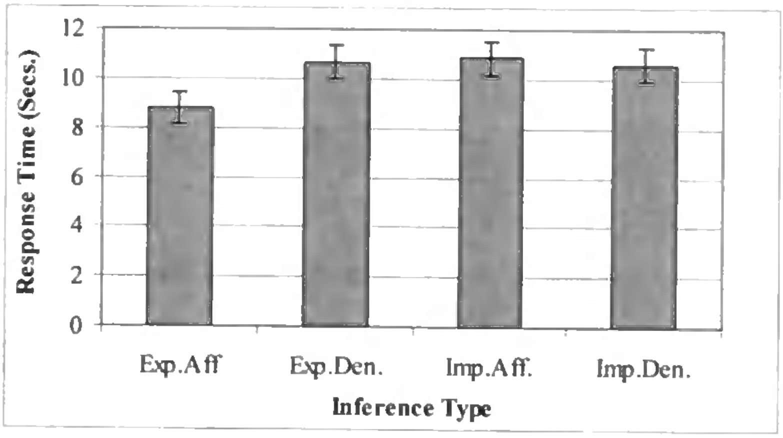
Results

Response Latencies

A 3x2x2 completely within subjects ANOVA was calculated using the response latencies, where the factors were *category pairing* (m-l, s-l and s-m), *reference* (explicit/implicit) and *inference* (affirmation/denial). As each participant received four inferences of each type (i.e. explicit affirmation, explicit denial, implicit affirmation, implicit denial), these were combined for the analysis, forming four response times for each of the three category pairings – 12 total response times per participant. By not including those times that were three standard deviations above, and two standard deviations below the mean, again controlled for outliers. This excluded six cases in which one or more response times fell outside the limit of 17-78 seconds.

The results showed two main effects of *referencing* ($F_{1,33} = 23.18$, $MSE = 6.751E7$, $p < .01$) and *inference* ($F_{1,33} = 17.44$, $MSE = 5.772E7$, $p < .01$). Explicit inferences took significantly less time to respond to than implicit inferences. Affirmation inferences also elicited significantly faster response times than implicit inferences. These two factors significantly interacted ($F_{1,33} = 21.75$, $MSE = 8.328E7$, $p < .01$) and this interaction is shown in Figure 5.8. It is apparent that the significant effects are almost entirely due to the faster response times on explicit affirmation type inferences, with the mean times for the other three inference types being very similar to each other.

Figure 5-8. Mean Response Times by Referencing and Inference (Exp.5).

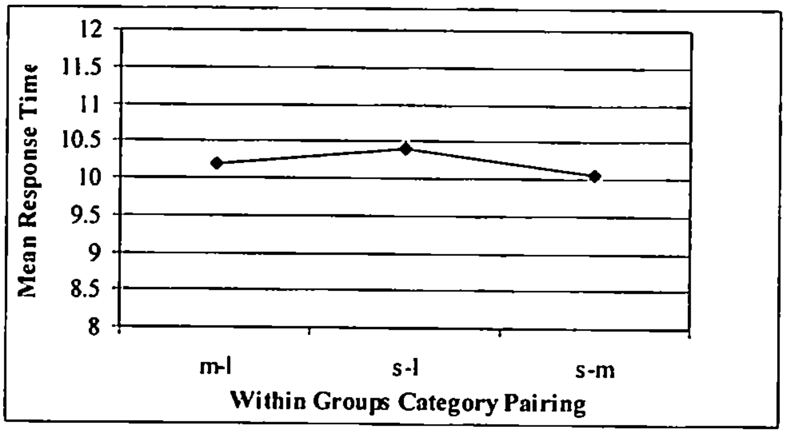


(NB Error bars indicate +/- 2 standard errors)

The response times differ from Experiment 4 where the pattern was of the form exp.aff>exp.den.>imp.aff>imp.den. Here the pattern takes the following form, exp.aff.>imp.den.>exp.den.>imp.aff. but it must be considered that the response times in the present study have not had the benefit of having the sentence verification times subtracted, as the sentence verification task was not included in the study. Nonetheless, it would seem that the only consistent outcome is that explicit affirmation problems elicit the fastest response times. The responses to the other three inference types are in both experiments, within a similar narrow time margin and it might well be expected that these three times do not follow a rigid pattern between experiments.

It would seem that the concept size manipulation did not have any significant effect on the overall response times as can be seen from Figure 5.9. As in the previous experiment, it was predicted that inferences made up from smaller concepts would affect response times in that they would be generally quicker to respond to than inferences that contained larger concepts.

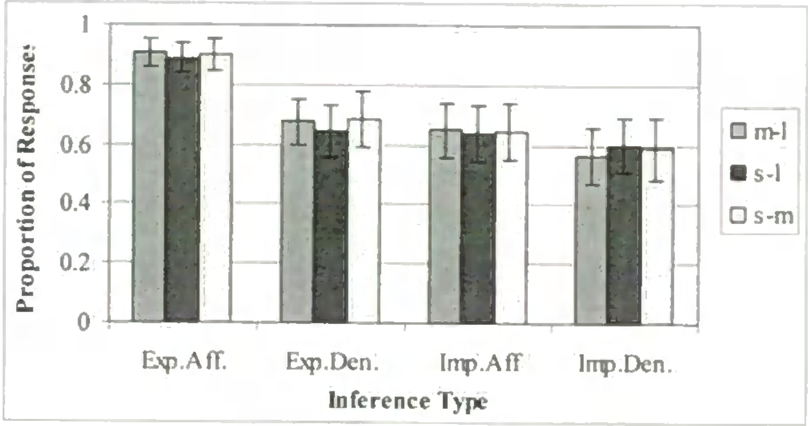
Figure 5-9. Mean Response Times by Category Pairing (Exp.5).



Responses

Figure 5.10 shows that the matching bias effect is occurring for all three category pairing conditions in that there are generally more 'follows' responses to explicit than implicit inferences. Compared to Experiment 4, there are generally more responses to all inference types apart from explicit affirmation inferences. It can be also seen that there is a greater matching effect on denial inferences than in Experiment 4, shown by a larger proportion of explicit denial inferences than implicit denial inferences. The same ANOVA design for the latencies was used for the responses which revealed two main effects of; *reference* ($F_{1,39}$ = 49.41, MSE = 1.107, $p < 0.01$) and *inference* ($F_{1,39}$ = 26.40, MSE = 1.547, $p < 0.01$) and a significant interaction between the two ($F_{1,39}$ = 12.28, MSE = 1.141, $p < 0.01$). It can be seen from Figure 5.10 that the proportion of responses for explicit affirmation inferences may be largely responsible for the significant findings, with the other three inference types eliciting similar amounts of 'follows' responses.

Figure 5-10. Proportion of Responses by Referencing, Inference and Category Pairing (Exp.5).



(NB Error bars indicate +/- 2 standard errors)

The percentage of 'follows' responses for each logical case for the 16 negations paradigm inferences can be seen in Table 5.6. Compared with Table 5.2 it can be seen that there are generally more 'follows' responses on denial inferences (DA and MT). It can only be assumed that either the materials used in the study encouraged more endorsements in some way or that differences occur between experiments due to population variance.

Table 5.6. Percentage of 'Follows' Responses for Each Logical Case in Each Category Pairing (Exp.5).

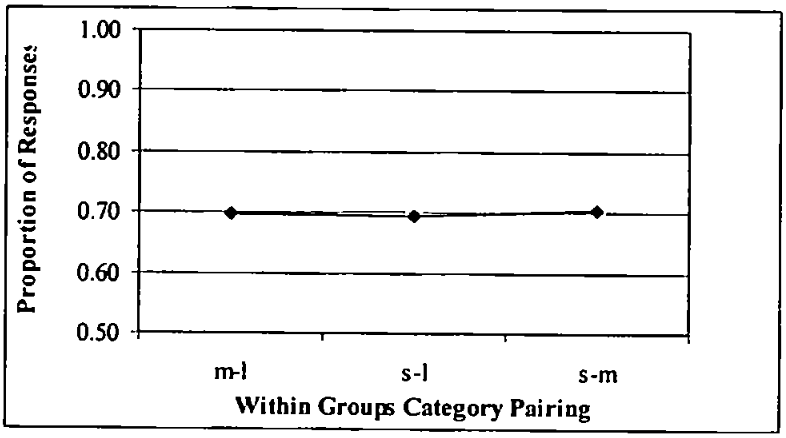
	M-l Category Pair				*S-l* Category Pair				*S-m* Category Pair			
	MP	DA	AC	MT	MP	DA	AC	MT	MP	DA	AC	MT
If P then Q	88	68	90	68	93	73	80	73	90	63	78	75
If P then not Q	98	53	55	80	93	58	60	83	95	40	55	78
If not P then Q	76	83	88	38	73	68	90	38	78	78	98	58
If not P then not Q	68	58	60	50	58	53	65	55	65	55	60	63

NB Matching cases underlined

Concept Size

With regard to the category size manipulation, there was no effect on the responses given by participants. This can be clearly seen from Figure 5.11. It was predicted that more 'follows' responses, and therefore more matching bias, would be elicited using *s-m* category pairings. This was not the case in that it would seem that matching bias occurred at the same level across all three pairs.

Figure 5-11. Proportion of Responses by Category Pairing (Exp.5).



API

The Affirmative Premise Indices were calculated as in Experiment 4. Table 5.7 shows the one-sample t-test values to determine whether the scores were significantly different from zero, which gives an indication as to whether a significant matching bias effect is present.

Table 5.7. One-sample t-test values for the API Scores for Each Category Pairing (Exp.5).

M-l Category Pairing	*S-l* Category Pairing	*S-m* Category Pairing
6.21 *	4.35 *	5.66 *

* *Denotes significance at* $p < 0.01$.

It can be seen that matching bias was significantly present in all three conditions. The present results do not replicate the previous finding (Exp.4) of a lack of a matching bias effect using larger sized concepts. In the present experiment, there would seem to be no difference in matching bias between the three category pairings.

NCI

Calculating the NCI as in Experiment 4 revealed significant effects of conclusion polarity using all three category-pairings. Table 5.8 shows the one-sample t-test results for each category pairing.

Table 5.8. One-sample t-test values for the NCI Scores for Each Category Pairing (Exp.5).

M-l Category Pairing	*S-l* Category Pairing	*S-m* Category Pairing
4.03*	3.83*	3.40*

* *Denotes significance at* $p < 0.01$.

The present results differ from those found previously in that there appears to be a negative conclusion bias in all three conditions, whereas this was not present in Experiment 4 using the smallest sized concepts. Again, the present experiment has failed to replicate the previous findings of a concept size effect.

Discussion

The findings show that there is a robust effect of matching bias. This has occurred at almost equal levels across all three conditions. Experiment 4 showed that category size had an effect on matching bias, but at a weak level. The present experiment has not even indicated a weak effect. The two main differences between the experiments are the totally within design and the use of more nested triple categories. It was seen in Experiment 4 that the strongest concept size effect occurred between groups. It could be that the larger proportion of inferences that participants had to respond to (48 compared to 32) might have induced an overall strategy of endorsement, which is seen by the larger proportion of responses generally. Factors such as boredom may have been playing a part. Also, the shear volume of varying concepts that were being activated in memory might have diminished differences in activation levels, spreading the trace activations across several semantic networks.

Concept Size

A more likely explanation for the lack of an effect with the category pairings is the selection of material that was used. There is no real way of controlling for the concept sizes unless participants see all three nested concepts. That way, the relative sizes can be accounted for i.e. *mineral* is a larger concept than *gem*, which is then a larger concept than *diamond.* In the present experiment, participants only saw two of the nested triple concepts, and it is arbitrary just how these pairs compare in size. Just because the concepts in nested triples vary in relative size (1<2<3), it does not necessarily mean that the concepts equal the size of concepts in other nested triples. For example, the triples *mineral/gem/diamond* and *structure/dwelling/house* have concepts decreasing in size relative to each other, but comparing the size of the concepts away from their nested triple is problematic. A *house* for instance may not necessarily be the same-sized concept as a *diamond.* Therefore, a concept's placement within an arbitrary triple does not guarantee its equivalent size with other concepts. It might have therefore been more prudent to use the *same* nested triples for each condition thus guaranteeing relative concept size differences. With hindsight it is felt that concept size was not correctly manipulated within the present study.

Summary

The two experiments have shown clear effects of matching bias in the conditional inference task using materials that do not lexically match. It was also shown that a matching bias effect may be occurring in the sentence verification task, by means of reduced response latencies. Due to the correlation between response latencies and correct responses it was suggested that matching terms in a sentence verification problem may have a facilitatory effect on cognitive processing. With regard to possible concept size effects, the verification task did not replicate previous categorisation reaction time findings such as Landauer and Freedman (1968).

In Experiment 4, the concept size manipulation showed a trend in inference endorsements in line with the prediction that there would be fewer matching responses to inferences constructed from larger rather than smaller concepts. The API measure of matching bias indicated that the concept size manipulation may be having a significant effect on matching bias in that it was no longer significantly present using the s-l-size category pairings in the *m-l/s-l* condition. The large-size category pairings did not replicate this finding though which posed problems for the hypothesis.

Due to possible carry-over effects using the same nested triple for each condition, a further inference task was created in Experiment 5, utilising more materials in a completely within design. This did not replicate any concept-size effects on matching bias. It was suggested that the experimental manipulation of concept-size was flawed and may have led to the lack of a replicated effect. It was pointed out that it is problematic using materials of this semantic nature to control absolutely for concept size and the proposed spread of mental activation that these concepts are suggested to induce. The conditional inference experiments in the following chapter seek to rectify some of these earlier problems and to clarify some of the interesting findings that have so far arisen. They also aim to investigate other issues with regard to negation, logical outcome and word familiarity.

Chapter 6

Experiments 6 & 7 – Conditional Inference Task

Experiment 6

For Experiment 6, concept size is manipulated but this time the same nested triples are used for each condition, thus assuring the relative concept size differences, as was the case in Experiment 4. In Experiment 4 it was felt that there might have been carry over effects, with participants seeing all three concepts from a nested triple. This was avoided by using a completely between-subjects design. As has been seen, the inference task usually elicits a strong matching bias effect when comparing responses to the four inference types of explicit/implicit affirmation/denial.

The disadvantage of comparing the four inference types, especially along the lines of explicit/implicit referencing in order to detect matching bias, is that there is no control over the presence or absence of negation in both the major premise and conclusion. For instance, the explicit affirmation *modus ponens* inference, '*If A then 2. A, therefore 2*' is being compared to the implicit affirmation *modus ponens* inference, '*If not A then 2. B, therefore 2*'. According to the mental activation model, the clause 'not A' will be activated at a lower level than the clause 'A'. Calculating matching bias by comparing the responses to both inferences is not controlling for the presence of negation in the categorical premise*referred clause pairing ('*A*A' - matching, 'B*not A' - mismatching*).

It was seen with the verification task in Experiment 4 that negation can cause differences in responses and latencies. It would be better to compare responses to inferences when there is '*A*A*' and '*B*A*'. Although there are inferences where this happens (e.g. denial of the antecedent: *if A then 2. B, therefore not 2*) logic is confounded in that a comparison of affirming MP inferences will be made with denying DA inferences. The polarity of the conclusion also changes. Conclusion polarity can affect responses as is seen by previous NCI results.

To compare matching and mismatching cases, but keeping the polarity of the premises constant would involve using explicit and implicit referencing in the categorical premise e.g. 'not A' and 'B'. This can only be done with what up until now have been labelled as *implicit* inferences, because only these inferences have a categorical premise that negates the referred clause. Explicit negation has not been used for the categorical premise in the tasks so far, as the explicit/implicit distinction has relied on the fact that half of the minor premises from the 16 negations paradigm inferences are explicitly affirmative, and half are implicitly negative. (Affirmative minor premises are naturally explicit e.g. 'if A then 2. *A*, therefore 2', and negative minor premises are naturally implicit e.g. 'if not A then 2. *B*, therefore 2'). In order to explicitly refer to the referred-clause in the latter instance would require substituting 'not A' for 'B'. No such substitution can be made for inferences with affirmative minor premises and so only MP3, MP4, AC2, AC4, DA1, DA2, MT1 and MT3 inferences can be used (see Table 3.1, Chapter 3). By using explicit/implicit negation in this way allows for direct comparison of matching/mismatching cases using exactly the same inference structure in terms of the presence of negations in both the major premise and conclusion (whilst not confounding logic).

To achieve this with categorical materials involves reversing the concept sizes between the referred clause and categorical premise. Previously, the referred clause has been constructed from the largest of the two concepts (e.g. If there is a *tree* on the left... There is an *oak* on the left). This format no longer works when using explicit negation in the categorical premise, i.e. 'If there is a *tree* on the left....There is not an *oak* on the left'. It is apparent that the use of categories in this instant creates a problem that is *indeterminate*. There could be a 'beech tree' on the left, which does not deny the major premise. The information therefore that there is 'not an oak' offers no logical indication as to whether it denies or affirms the clause 'there is a tree'.

The concept sizes therefore have to be reversed in that the smallest category is now used for the referred clause and the largest for the categorical premise. For example, the inference, 'If there is an *oak* on the left....' can now be refuted by the minor premise, 'There is not a *tree* on the left'. Eight nested triples were used in this way creating 16 *determinate* problems i.e. there is a logical answer to the inference. Explicit and implicit negation were used for each of the eight inference types; MP3, MP4, AC2, AC4, DA1, DA2, MT1 and MT3 forming the 16 inferences in total.

Indeterminate Problems

Additionally, a further 16 inferences were also created which were classed as *indeterminate* i.e. there is no logical answer given the inference. An example of an indeterminate inference would be, 'If there is not an adder on the left....There is a reptile on the left'. The clause 'reptile' no longer affirms or denies the clause 'not an adder' as a reptile could be an adder or it could not. The indeterminate problems share the same structure as the eight inference types listed above. The only difference is the presence or absence of an explicit negation. So for instance, a determinate modus ponens inference could be, '*If there is not an adder on the left...There is not a reptile on the left*'. Whereas, the indeterminate inference of the same structure would not include the negation in the minor premise e.g. '*there is a reptile*' thus creating a problem without a logical solution.

By creating a set of *indeterminate* problems with both matching and mismatching categorical premise/referred clause pairings, allows the opportunity to study matching bias in a unique way. Usually, with the 16 negations paradigm inferences, half of the inferences *require* a 'follows' response because the inference conclusion is correct. Therefore, the usual measure of matching bias with determinate inferences also incorporates an amount of logical responding. With indeterminate problems, this is no longer the case. This should produce a clear picture of matching bias away from extraneous influences that can affect the amount of 'follows' responses such as the presence of explicit negations and logically correct answers. The prediction for *determinate* inferences remains the same as the previous experiments i.e. less matching bias will be seen using larger concepts. For indeterminate inferences, category size is predicted to affect matching responses although the direction of this effect is uncertain.

Participants

One hundred and eighteen undergraduate psychology students at the University of Plymouth participated in the experiment as part of their course requirement. Participants were tested in small groups of up to six at a time. Participants were randomly allocated to either of three conditions forming three independent groups for analysis, the numbers for which are; *m-l* concept-size (n = 28), *s-l* concept-size (n= 45), *s-m* concept-size (n = 45). The unequal participant numbers were caused by an anomaly with the computer randomisation process, but it was felt that there were still sufficient numbers in the *m-l* group to provide enough power for the experiment. The design was a 3x2x2x2 mixed design, with the three concept sizes being tested between groups (m-l, s-l, s-m), whilst *matching* (matching/mismatching), *inference* (affirmation/denial), and *logic* (determinate/indeterminate) being tested within groups.

The terms *matching/mismatching* were used here instead of the previous terms of *explicit/implicit*. The reason for this is that the terms explicit and implicit imply that the categorical premise either explicitly or implicitly *represents* the referred clause item. In this experiment, half of the inferences were *indeterminate* – that is they had no logical outcome. For example, given the premise 'If there is not an adder on the left...' the indeterminate categorical premise might be 'There is an animal on the left'. The categorical premise (an animal) no longer explicitly *represents* the referred clause (not an adder) because an animal could be an adder or it could not - but the actual items *match* in terms that an adder is an animal and that they are semantically related. *Four* each of the above, two-level, problem types were given to every participant, creating 32 problems in total, examples of which can be seen in Table 6.1.

Table 6.1. Examples of the Matching Determinate and Indeterminate Problems Used for Each Condition in Experiment 6.

	M-l Category Pairing (medium*large)[1]	*S-l* Category Pairing (small*large)[1]	*S-m* Category Paring (small*medium)[1]
Determinate Inference	- If there is NOT a *reptile* on the left then there is a *gem* on the right. - There is NOT an *animal* on the left. - Therefore there is a *gem* on the right.	- If there is NOT an *adder* on the left then there is a *diamond* on the right. - There is NOT an *animal* on the left. - Therefore there is a *diamond* on the right.	- If there is NOT an *adder* on the left then there is a *diamond* on the right. - There is NOT a *reptile* on the left. - Therefore there is a *diamond* on the right.
Indeterminate Inference	- If there is NOT a *reptile* on the left then there is a *gem* on the right. - There is an *animal* on the left. - Therefore there is a *gem* on the right.	- If there is NOT an *adder* on the left then there is a *diamond* on the right. - There is an *animal* on the left. - Therefore there is a *diamond* on the right.	- If there is NOT an *adder* on the left then there is a *diamond* on the right. - There is a *reptile* on the left. - Therefore there is a *diamond* on the right.

[1] Concept size for the referred-clause*categorical premise shown in parentheses

Design and Procedure

Category pairings were created using nested triple superordinate categories and subcategories, for instance, 'animal', 'reptile' and 'adder' – decreasing in concept size from large, medium to small. Pairs taken from each of the nested triples were used to create the three groups: *M-l* category pairing (e.g. reptile & animal), *s-l* category pairing (e.g. adder & animal), *s-m* category pairing (e.g. adder & reptile). These pairings were then placed in the conditional problem formats, examples of which can be seen in Table 6.1.

Eight nested triples (see Appendix F) were used to form the 32 problems. The allocation of these concepts to the 32 different conditional inferences was randomised.

Instructions

At the start of the experiment participants were given a sheet of paper with the following instructions:

> *The experiment involves a series of 32 problems. Each problem consists of a RULE concerning a set of fictitious cards. All of the 32 rules concern pictures that appear on sets of cards. You are to IMAGINE that the cards have two pictures printed on them - one picture on the left side of the card and one picture on the right side of the card. There are ONLY two pictures on each card - one on the left and one on the right. The rule that is given states which pictures can appear together on a card.*
>
> *Immediately following each rule you will be given a STATEMENT about a picture that appears on one of the cards. Your task is to decide whether the CONCLUSION that is given about the card actually 'follows' or 'does not follow' the rule in light of the statement that was given.*
>
> *An example would be:*
>
> *(RULE)   If there is a cat on the left side, then there is a book on the right side.*
>
> *(STATEMENT)   There is a cat on the left side.*
>
> *(CONCLUSION)   Therefore there is a book on the right side.*
>
> *(Answer: FOLLOWS)*
>
> *Use the marked keys on the keyboard to make your response.*
> *The blue key on the left is for a 'follows' response.*
> *The red key on the right is for a 'does not follow' response.*
> *Each screen will show these two options so you don't have to remember them!*
>
> *The problem number will appear at the bottom of the screen so you know how many you have done.*
>
> *Try to respond as quickly as possible but avoiding mistakes. When you are ready, press the mouse button to continue.*

Only two keys were programmed for a response; the 'Z' key on the keyboard was coloured blue, and the '2' on the number keypad was coloured red. Participants always had a reminder of the appropriate key for the response they might wish to make, from the colour and position of the *'follows/does not follow'* words at the bottom of each screen and the corresponding position/colour of the keys on the keyboard.

Once a response had been made to a problem, the screen cleared and there was a slight pause (approx. 2 secs) before the next problem was displayed. Latencies were timed from the display onset of the problem, and ceased as soon as a response was made.

Results

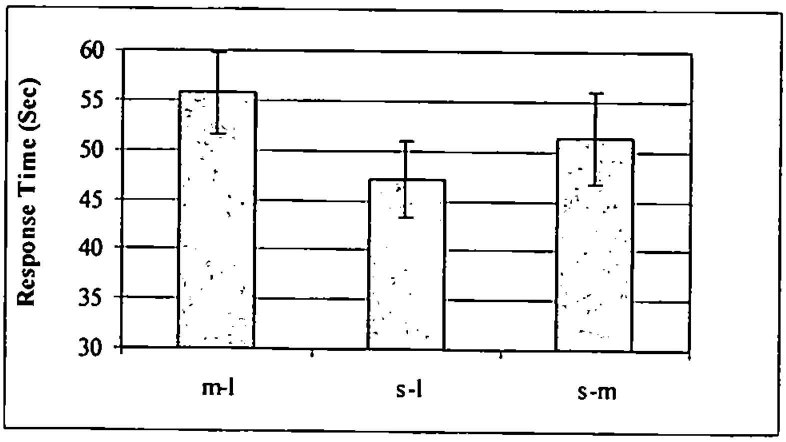
It should be noted that the following analysis includes both indeterminate problems and determinate ones, and strictly speaking, the indeterminate problems do not provide true *affirmation* and *denial* problem types. The indeterminate problems have been coded in this way though, in order to differentiate the two different problem structures. An example will hopefully clarify this situation. Looking at only the referred clause and inferential clause in the conditional inference problems that were used, it can be seen that in *determinate* explicit-affirmation problems, the referred clause might be, 'NOT an adder' and the categorical premise 'NOT a reptile'. The *indeterminate* equivalent, which is also going to be labelled as 'explicit affirmation', would have 'adder' as the referred clause and 'reptile' as the categorical premise. Logically, the concept of 'reptile' does not necessarily represent 'adder' and can not therefore be classed as *affirming*. It would on the other hand be affirming if the two concepts were reversed within the problem so that 'adder' would then represent 'reptile'.

Response Latencies

A 3x2x2x2 mixed ANOVA was calculated, with the three category pairings being tested between groups (m-l, s-l and s-m), whilst *matching* (matching/mismatching), *inference* (affirmation/denial), and *logic* (determinate/indeterminate) being tested within groups. Response latencies were combined forming eight response times per participant, along the three factors of determinate/indeterminate, matching/mismatching, affirmation and denial. As pointed out previously, the indeterminate problems were neither affirmation or denial problems, but were classified as such when they represented the corresponding problem structure – only on determinate inferences is the classification of any real relevance but it enabled comparisons of problem structures with indeterminate problems.

Excluding times that were two standard deviations below, and three standard deviations above, the overall mean, controlled for outliers. This excluded 19 cases in total (6 from the *m-l* group, 5 from the *s-l* group, 8 from *the s-m* group) where one or more total times fell outside of the limit of 6-123 sec. The results showed a main effect of category-pairing, $F_{2,97} = 3.51$, $MSE = 1.235E9$, $p < .034$. The mean times are shown in Figure 6.1.

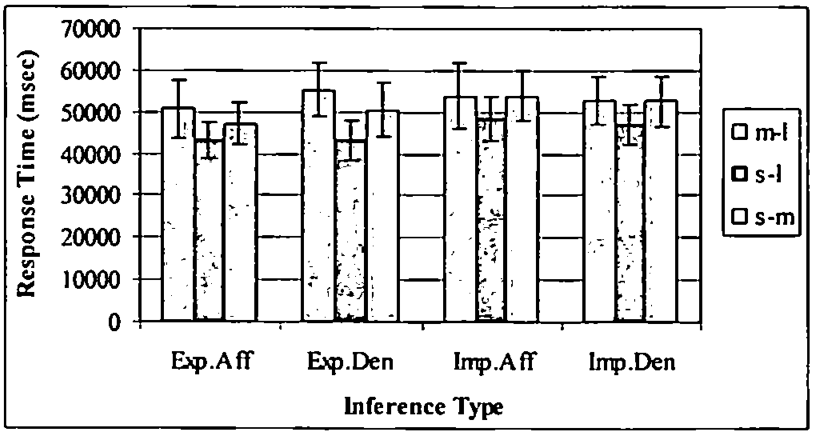
Figure 6-1. Mean Combined Response Times for the Three Category Pairings (Exp.6).



(NB Error bars indicate +/- 2 standard deviations)

Using a post-hoc LSD follow-up analysis revealed that there was a significant difference between the *m-l* and *s-l* groups response times ($p = .011$). It would appear that the smaller concepts used in the *s-l* group as opposed to the concepts used in the *m-l* group produced faster response latencies when encountering the inference problems. This effect can be seen across all four inference types as seen in Figure 6.2.

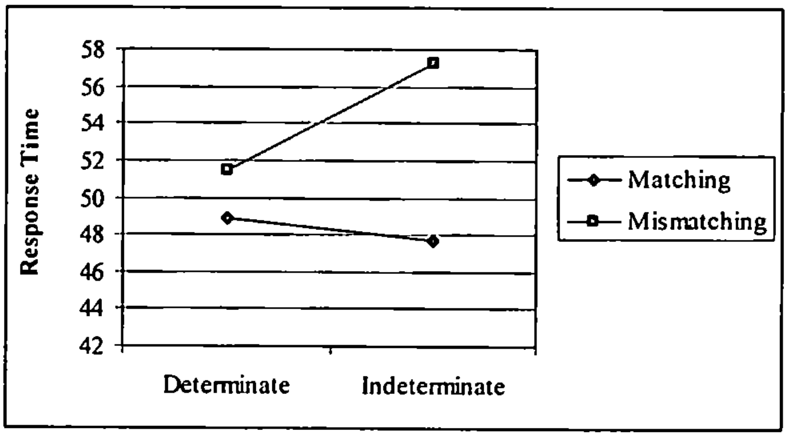
<u>Figure 6-2.</u> Average Response Latencies for Determinate Inferences for the Three Category Pairings (Exp.6).



(NB Error bars indicate +/- 2 standard deviations)

There was a significant main effect of *matching* ($F_{1,97} = 29.08$, $MSE = 2.315E8$, $p < .01$) in that matching inferences produced faster response times (6 sec on average). This effect was greater for the indeterminate problems, as can be seen by the interaction in Figure 6.3 ($F_{1,97} = 16.34$, $MSE = 1.449E8$, $p < .0001$).

<u>Figure 6-3.</u> Interaction of Matching and Logic on Response Latencies (Exp.6).



The greatest difference between matching and mismatching response times occurs for the indeterminate problems. This pattern is particularly interesting when compared to the actual responses that were given, as will be seen in the next section on the analysis of 'follows' responses. Although matching terms produce faster response latencies, it will be seen that this does not necessarily follow the pattern of responses that are actually given i.e. a faster response does not mean that the inference is actually endorsed.

Responses

The proportion of responses by inference type for both matching and mismatching inferences are shown in Table 6.2.

Table 6.2. Proportion of 'Follows' Responses For Each Logical Case.

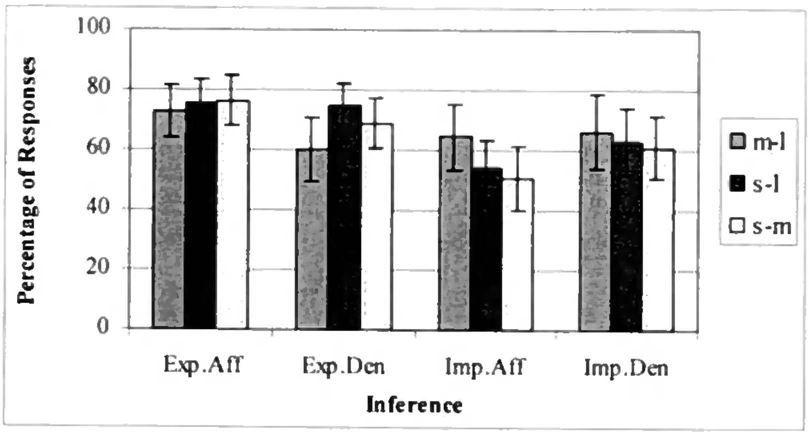
Determinate	M-l				S-l				S-m			
	MP	DA	AC	MT	MP	DA	AC	MT	MP	DA	AC	MT
If P then Q		79		68		87		93		84		71
		79		75		69		67		73		62
If P then not Q		39	61			62	67			67	71	
		54	75			60	64			58	56	
If not P then Q	79			54	78			58	80			53
	75			57	71			56	62			51
If not P then not Q	68		82		87		69		76		78	
	54		54		38		42		47		38	

Indeterminate	M-l				S-l				S-m			
	MP	DA	AC	MT	MP	DA	AC	MT	MP	DA	AC	MT
If P then Q		46		14		9		16		18		4
		32		32		42		22		33		40
If P then not Q		25	32			24	22			11	18	
		36	36			27	44			38	31	
If not P then Q	32			25	20			18	22			13
	29			36	31			40	33			24
If not P then not Q	39		29		27		18		22		29	
	25		29		24		24		20		27	

* Endorsements of inferences with matching (implicit) categorical premises underlined.

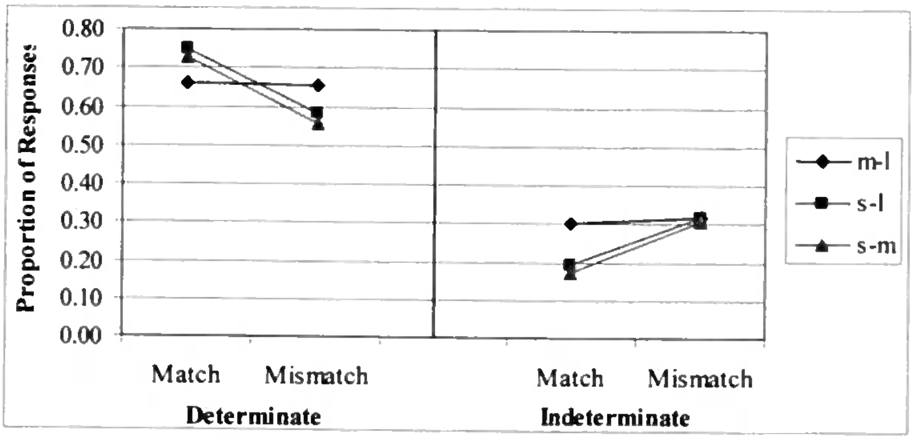
Figure 6.4 shows the proportion of responses collapsed into the four inference types. The endorsement responses ('follows') were analysed using the same ANOVA design as for the response latencies. Differences in matching/mismatching responses can be detected by differences in the amount of endorsements. The total amount of 'follows' responses were calculated along the same criteria as the latencies – providing eight scores per participant, ranging from 0-4. The analysis showed a strong main effect of *logic* ($F_{1,115} = 352.12$, $MSE = 1.533$, $p<.01$) which can be seen in Figure 6.5 where the determinate inferences elicited far more 'follows' responses. The analysis also showed a three-way interaction between category pairing, matching and logic ($F_{2,115} = 4.55$, $MSE = 1.434$, $p = 0.013$) which is seen in Figure 6.5

Figure 6-4. Percentage of Responses for Determinate Inferences for All Three Category Pairings (Exp.6).



(NB Error bars indicate +/- 2 standard deviations)

Figure 6-5. Interaction between Category Pairing, Matching and Logic 'Follows' Responses (Exp.6).



The analysis highlights two interesting phenomena with the materials. Firstly, there seems to be a matching bias effect with the materials used in both the *s-l* and *s-m* groups for determinate problems. This effect is *reversed* for indeterminate problems in that there are fewer matching 'follows' responses than mismatching ones. Secondly, when large concepts are used, there appears to be no matching bias effect at all.

Matching

A further analysis was carried out on the *differences* between matching and mismatching responses i.e. a score of matching minus mismatching responses was calculated for each participant. This measure is essentially the same as the API scores, in that previously, matching minor premises have also been affirmative ones. In the present study though, all the minor premises were negative, but as before with the API, responses to mismatching inferences were subtracted from responses to matching inferences. A two-tailed *t*-test was carried out on these difference measures to see if they were significantly different from zero. A significant result would indicate that there is significantly more 'follows' responses being made to matching inferences. The results can be seen in Table 6.3.

Table 6.3. One-Sample t-test Values on the Difference in the Amount of Endorsements for Inferences with Matching and Mismatching Referred Clause/Categorical premise Pairings (Exp.6).

	M-l	S-l	S-m
Determinate	0.09	3.40 *	3.35 *
Indeterminate	-0.26	-2.79 *	-3.93 *

**Denotes significance different from zero at p < 0.01.*

It is apparent from Table 6.3 that a significant amount of matching is taking place for determinate inferences in the *s-l* and *s-m* groups, but not in the *m-l* group. The *t*-test values for indeterminate inferences also indicate that there is a *mismatching* phenomenon occurring in the *s-l* and *s-m* groups, but again not in the *m-l* group. It would seem that the category pairings used in the *m-l* condition have eradicated any differences between matching and mismatching materials for both determinate and indeterminate problems. A one-way ANOVA on the difference of endorsements for matching and mismatching inferences for determinate inferences revealed an almost significant difference between conditions ($F_{2,115} = 2.84$, $MSE = 1.580$, $p = .062$).

NCI

As in the previous two experiments, the NCI scores were calculated for each participant. These are shown in Table 6.4.

<u>**Table 6.4.**</u> One-sample t-test values for the NCI Scores for Each Category Pairing (Exp.6).

	M-l	S-l	S-m
Determinate	1.61	1.21	1.13
Indeterminate	-0.10	-1.95	0.09

No t-values reached significance at $p < .05$

The t-test on the NCI scores for the determinate inferences in the *m-l* condition revealed a $t = 1.61$, $p = .059$, indicating a near significant amount of negative conclusion bias taking place. This finding almost replicates the NCI results in Experiment 4, where only the inferences using larger concepts elicited negative conclusion bias. Comparing the API and NCI results, it appears that when the API scores are significantly different from zero, the corresponding NCI scores are not significant. The concepts used in the *m-l* group appear to be responsible for these differences in that they do not elicit matching bias between the categorical premise and referred clause, but produce heightened negative conclusion bias. This replicated finding offers strong evidence that larger concepts reduce matching bias, as predicted by the mental activation model. At present, no other account of matching bias can explain these findings. What is as yet unexplained is why reduced matching bias is found together with significant negative conclusion bias.

Discussion

The response latency results showed significant differences between the *m-l* and the *s-l* groups in that overall, the inferences in the *m-l* group took longer to respond to. It would appear that the use of larger concepts might be creating a greater degree of problem difficulty than the use of smaller concepts. The pattern of response latencies (exp aff < exp den < imp den < imp aff) inversely follow the proportions of actual responses (exp aff > exp den > imp den > imp aff). The smaller the amounts of time taken to respond to an inference type, the more endorsements that are given.

If latency and amount of endorsements are an indicator of problem difficulty, then it poses the question as to what makes a problem appear more difficult than other problems. Half of the inferences (in each of the four determinate inference types) are logically fallacious and if reasoners were responding logically, then it should be expected that for each of the four types, there should be a 50% endorsement rate. Generally the endorsement rate is much higher than this, and it would appear that logical requirement of the task is not connected to problem difficulty. It seems more likely that problem structure and content affect response latencies and the amount of endorsements given. The main factor in this experiment that seems to exert the greatest influence on both latencies and the amount of endorsements is the explicit (matching)/implicit (mismatching) referencing structure of the inference. The explicit/implicit referencing distinction is basically another term for categorical premise/referred clause matching. So it would appear that 'matching' is responsible for perceived problem difficulty.

The *m-l* group effectively shows that there was not a matching effect between the larger concepts used for the categorical premise and referred clause in these inferences. The *m-l* group also elicited the slowest response latencies overall. There appear to be two levels at which matching bias is taking place within this experiment. The first level is the explicit/implicit referencing taking place within each group of inferences. Matching bias is seen here for explicit inferences where the referred clause and categorical premise items match rather than mismatch. The second level at which the matching bias effect can be seen is when concept size is manipulated. The larger the concept size, the less of a matching bias effect.

The mental activation model describes how the activation of large concepts in memory requires a larger spread of the activation power that is available. This spread of activation reduces the activation *level* caused by 'overlapping' concepts and therefore the concepts appear as less relevant (reduced matching effect). Evans' (1998) account of matching bias in terms of relevance is pertinent to the current findings in that it would appear that larger concepts appear less relevant and create what would appear to be greater difficulty in choosing a response to the problem. If the items in the inference do not appear relevant, then processing the problem seems to become more difficult which is seen by fewer endorsements and longer response latencies. It would appear that the terms; relevance, problem difficulty and matching bias may all be the descriptions of the same phenomenon, and the results of the concept size manipulation here offers support for the neural network/semantic association approach formulated in the mental activation model as an explanation of this phenomenon

By using the same inference structure but manipulating the categorical premise clause in terms of explicit or implicit referencing has enabled a more controlled view of matching bias effects. Across all four inference-types, there is a more even amount of responses for all four inference types, which is seen in Figure 6.4. This is especially apparent in the reduced amount of explicit affirmation inference endorsements compared with the two inference task experiments in Chapter 5. It would seem that by controlling for the presence or absence of negations in the major premises and conclusions has controlled for other factors that may have previously led to an increase or decrease in inference endorsements.

With regard to the responses to indeterminate problems, they have revealed an unusual effect of matching bias in that reasoners seem *less* likely to endorse the inference in which the categorical premise/referred clause terms match. This finding sheds new light on the matching phenomenon. For the determinate problems, it was pointed out that half of them are logically fallacious, specifically those falling under the categories of 'denial of the antecedent' (DA) and 'affirmation of the consequent' (AC). (Although, with a biconditional interpretation of the conditional premise, endorsing these inferences could be seen as a correct response). As more matching inferences are endorsed than mismatching ones, it could be argued that matching terms with determinate instances facilitates processing, and therefore elicits more follows responses (a similar facilitatory effect was pointed out in Exp. 4 using the sentence verification task).

If matching terms facilitate processing the inference then it can be seen that for indeterminate inferences, matching terms facilitate giving the correct response, which in this case is the 'does not follow' response. It would therefore not be surprising then that matching terms for indeterminate inferences elicit fewer endorsements ('follows' responses). What is being seen in the matching results for determinate and indeterminate problems is not just a 'blind' automated response to matching items but an indication that matching terms actually *facilitate* the processing of the problem. Reasoners aren't just *endorsing* an inference because they see it as relevant. Instead, matching items can actually aid in the refutation of an inference. This is a new finding in terms of the matching bias phenomenon and indicates strongly that System2 conscious thought processes are working simultaneously with the System1 representation stage.

Experiment 7

The conditional inference studies so far have raised three main issues with regard to the effect of concept size and matching bias. These concern the following; the general *comprehension* of the concepts used in the problem; the notion that a heightened *perception of relevance* is reflected through a specific response choice; and the nature of the *matching pair relationship*. It will be argued that the first two issues can still be incorporated within the proposed mental activation model, but the latter issue of the *nature of the relationship* between matching items needs further experimental clarification.

The previous chapter indicated that concept size could affect performance on the conditional inference task. It was seen that when large categories were used as concepts in a conditional inference, response times increased together with a subsequent drop in the proportion of endorsements for those inferences. Experiments 4 and 6 have showed a pattern in the proportion of inference endorsements ('follows' responses) in the direction predicted by the mental activation model to the extent that the use of larger concepts reduces matching bias. More matching bias was taking place in the group where participants were given inferences with smaller sized concepts.

Comprehension

It is possible that the differences in response patterns which occurred by manipulating concept size may have been due to simple *comprehension* difficulties, rather than overlapping trace activations in memory of matching terms proposed by the mental activation model. The mental activation model does not take into account differences in response patterns when mismatching concepts are used in the inference. If two concepts do not initiate an overlapping pattern of activation then they would be perceived as equally irrelevant, irrespective of the size of the concepts.

If difficulty in the comprehension of larger sized concepts was responsible for differences in response rates, the previous conditional inference experiments would show an *overall* decrease in response rates for inferences that were constructed from large concepts. This effect would be expected for inferences that had *both* matching and mismatching terms because both types of inference would still have contained large concepts. Experiments 5 and 6 in particular did not show any substantial overall differences in response rates between conditions. Experiment 6 shows quite clearly that differences in response rates between groups occurred with the matching indices. The inferences utilising large concepts did not show any significant signs of matching bias, yet overall the response rates were similar to the other two groups.

Perceived Relevance

With regard to *perceived relevance* being shown by the amount of endorsement to the inference, the results from Experiment 6 suggest that the picture is more complex. The facilitatory effects of matching materials in *refuting* the inference (as was the case for the indeterminate type problems) suggests more than just a 'follows' response as being an indicator of the inference being perceived as relevant. With the *indeterminate* problems, it was found that matching items made the 'does not follow' response appear more *relevant* when using smaller concept sizes (this effect was not seen using larger sized concepts). It could be then that the smaller concepts facilitated the processing of the problem. If a biconditional interpretation of the conditional problem is taking place, then it can be assumed that matching with smaller concepts aids in the correct logical response to the problem. Larger concepts somehow interfere with this process.

The proposed mental activation model only puts forward an explanation for the notion of a heightened perception of relevance to matching materials in terms of activation overlap. The indeterminate problems in Experiment 6 may show that heightened perceptions of relevance do not necessarily mean that an endorsement of the inference will take place. It could quite simply be the case that either of the two possible responses ('follows' or 'does not follow') could be seen as relevant, depending on the perceived logical requirement of the problem. Up until now, previous studies have taken for granted that matching bias is observed by an *increase* in the amount of endorsements of the problems. Thus, if the problem materials were perceived as relevant, then it was assumed that reasoners would automatically endorse the inference.

Matching Relationship

The final point here concerns the *nature of the relationship* between the matching concepts. The issue is again raised (as in Chapter 4) as to what is the precise nature of the 'matching' connection between the concept used for the categorical premise and the concept used for the referred clause. Is it purely the case that a larger concept has more neural connections in memory and the activation of such a concept is at a lower *level* of activation because of the spread of interconnected concepts? Previous studies concerning concept size and memory were outlined in Chapter 3 where it was pointed out that concept size alone might not account for differences found in response times using simple sentence verification tasks. Issues such as word relatedness (measured against production frequency norms) and word familiarity were raised. It could be that the effects found in the conditional inference tasks so far might not necessarily be down to concept size but one or more of these other factors.

There is the possibility that the materials may be affecting *processing* of the problem in terms of the familiarity or relatedness of the concepts that are used. This issue will now be addressed in the present experiment. It could be that the concepts 'tree' and 'oak' are more familiar or related than 'tree' and 'plant'. Familiarity or relatedness of the matching items would of course only affect responses given to inferences that contained matching categorical premise/referred clause items and could therefore offer an alternative explanation of the matching bias findings so far. It was seen in Experiment 6 that using larger sized concepts completely eradicated the matching bias effect. The proportions of endorsements were almost equivalent for inferences with matching and mismatching cases. This result could be suggestive of the lack of association or familiarity between the two matching terms such that they appear as equally dissociated. For instance, reasoners may fail to make the connection that 'a tree' represents 'a plant' (matching), just as they would not make the connection that 'a gem' represents 'not a plant' (mismatching). Or, if they do make a connection between matching terms, then at best it may be at a far-reduced level than had they been given the terms 'oak' and 'tree' for instance.

Experiment 7 sets out to discover by manipulating the *relatedness* of the two matching terms (categorical premise and referred clause), whether similar results can be achieved as were found by manipulating the size of the concepts. Such a finding would pose difficulties for the proposed mental activation model, which does not take into account the general frequency of usage or familiarity of particular concepts. The model concentrates solely on the initial activation of memory that occurs when encountering the concept in relation to the given problem. It does not make a direct account for previous familiarity or relatedness of the concepts in everyday use, but rather just how memory organises the concepts in a semantically structured network.

As seen in Chapter 3, Smith, Shoben and Rips (1974) claimed that previous studies into category size might have confounded the *relatedness* of the paired items. In their study, they controlled for relatedness by using 'production frequency norms'. These are formed by the amount of instances a group of participants produce to a prescribed superordinate category. An item would be described as being of frequent usage if it were produced the most number of times by the group of participants. For instance, given the superordinate category of 'snake', the instance 'adder' would be deemed as having more frequent usage than 'copperhead', if more participants produced it.

McCloskey (1980) claimed that Smith et al (1974) had confounded familiarity with relatedness, and showed that if participants had prior exposure to the materials, the relatedness effect diminished. It is not the current scope of this study to delineate the two issues, suffice to say that it is important to discover whether category size is mainly responsible for the previous findings in this study or whether theses extraneous variables are primarily involved. McCloskey's 'familiarity' findings concerned mainly the comprehension and response times to the sentence verification type tasks that were being used. Whereas in this study, what is of major interest is the actual response, as well as the response latency. Comprehension or reading times should therefore not affect the actual response given.

In this study, Brown's (1972) word norms were incorporated in which one of the most and one of the least produced category inclusions were used together with the superordinate category, to form the matching paired items. Therefore two conditions could be compared; one where the matching items consisted of a category and a frequently produced category member, and one condition where the same category and an infrequent category member were used. If familiarity or relatedness were factors rather than concept size, then there should be found to be a similar difference in response rates to the inference problems between groups as was seen by manipulating concept size.

Participants

Sixty-four undergraduate psychology students at the University of Plymouth participated in the experiment, either as part of their course requirement or for payment. Participants were tested in small groups of up to six at a time. Participants were randomly allocated to either of two groups: 32 participants in the *infrequent* group and 32 participants in the *frequent* group.

Design and Procedure

The design was a 2x2x2 factorial design, with level of *word frequency* being tested between groups (frequent/infrequent), whilst *referencing* (explicit/implicit), and *inference* (affirmation/denial) being tested within groups. Interactions of four of each of the last two 'within' factors were given to each participant creating the 16 problems necessary for the complete negations paradigm (i.e. 4x explicit affirmation, 4x explicit denial, 4x implicit affirmation, 4x implicit denial).

Using the standard conditional inference paradigm, the pairings of the categorical premise and the referred clause were manipulated along the lines of the level of frequency of the categorical premise in relation to the category represented by the referred clause. Eight pairs of instances were used for both infrequent and frequent groups (see Appendix G). So, as before in Experiment 6, the materials involved a category and category member, but this time, rather than manipulating category size with this pairing – it was the frequency of production of the category member that was manipulated. The level of production frequency (i.e. high or low) was established using Brown's (1972) word norms. These were originally constructed by asking 200 participants to produce as many items as possible in one minute that belonged to a given category (28 categories in total). The production frequency for lists of category members could then be produced, listing the most frequent to least frequent words generated. This was simply in the form of the amount of times a particular word was generated by the 200 participants. One example would be where the superordinate category of 'four-footed animal' was used, in which 'dog' was generated 189 times (most frequent) and several animals such as 'aardvark' were only generated by one of the 200 participants (least frequent). These frequency ratings were therefore used to generate the categorical premise items for the two conditions: high frequency (*frequent*) and low frequency (*infrequent*).

The conditional inferences were displayed on a computer screen as in the previous conditional inference experiments. Again, this involved the three lines of the conditional inference laid out as follows:

(*Infrequent* explicit affirmation premise - *frequent* concept in parenthesis)

> If there is a musical instrument on the left then there is a sport on the right.
>
> There is archery (football) on the right.
>
> Therefore there is a musical instrument on the left.

Participants' response latencies from the presentation of the inference and the actual response of either 'follows' or 'does not follow' were recorded.

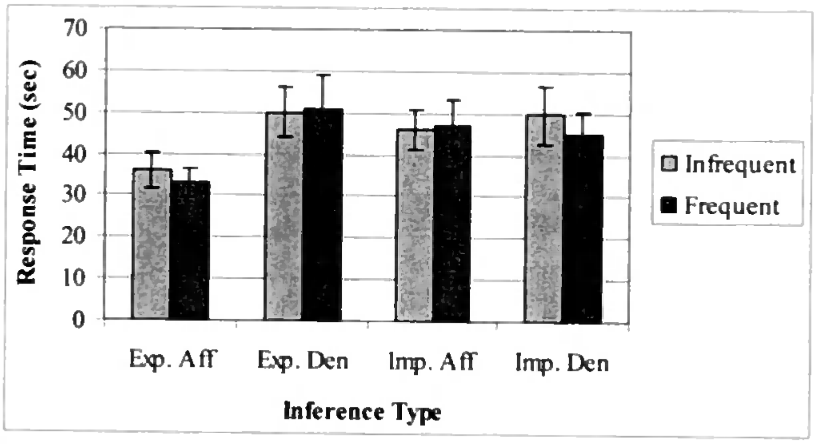
Results

Response Latencies

As there were four of each type of inference (i.e. explicit affirmation, explicit denial, implicit affirmation, implicit denial) allocated to each participant, the four latencies for each were combined. Outliers were controlled for by excluding combined times which were less than two standard deviations below the overall mean (11.3 sec) or higher than three standard deviations above the mean (100.5 sec). This excluded two cases from the *infrequent* group and three cases from the *frequent* group. A 2x2x2 mixed ANOVA was carried out on the combined latencies, with the factors being *word frequency* (infrequent/frequent), *referencing* (explicit/implicit) and *inference* (affirmation/denial).

The results showed significant main effects of *referencing* ($F_{1,57} = 9.33$, $MSE = 1.25E8$, $p < 0.01$) and *inference* ($F_{1,57} = 36.49$, $MSE = 1.16E8$, $p < 0.01$) and a significant interaction between the two ($F_{1,57} = 28.20$, $MSE = 1.21E8$, $p < 0.01$). There were no significant effects of *word frequency*. It can be seen from Figure 6.6 that the combined response times were very similar between infrequent and frequent words for each inference type. It would appear that the reduced response times to explicit affirmation problems is a major contributor towards the significant effects of *referencing* and *inference* type. The other three inference types seem to elicit similar combined response latencies (approx. 45-50 sec) irrespective of the frequency of the concept used for the categorical premise.

Figure 6-6. Average Combined Response Latencies for Each Inference Type in Each Frequency Group (Exp.7).

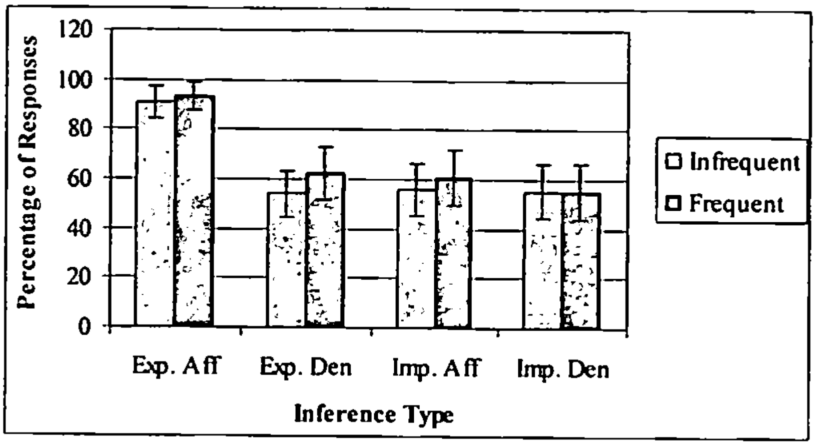


(NB Error bars indicate +/- 2 standard deviations)

Responses

The same ANOVA structure as that used for the response latencies was implemented. The combined amount of 'follows' responses were calculated for the four inference types. There was found to be significant main effects of *referencing* ($F_{1,62}$ = 22.36, MSE = 1.48, p< 0.01) and *inference* ($F_{1,62}$ = 39.29, MSE = 0.88, p< 0.01) and again, an interaction between the two ($F_{1,62}$ = 34.70, MSE = 0.68, p< 0.01). There were no significant effects of *word frequency* on the amount of inference endorsements ('follows' responses). Figure 6.7 shows that again, the main contributor to the significant effects is the large amount of 'follows' responses to explicit affirmation type problems. The other three inference types elicited almost equal proportions of endorsements.

Figure 6-7. Percentage of 'Follows' Responses for Each Inference Type in Each Frequency Group (Exp.7).



(NB Error bars indicate +/- 2 standard deviations)

Table 6.5 shows the breakdown of the proportion of endorsements for each of the 16 negations paradigm inferences. Overall there are equivalent amounts of endorsements in both conditions to inferences with matching and mismatching concepts. That is, there appears to be no difference in responses with words that have been classified as more frequent rather than infrequent.

Table 6.5. Percentage of Follows Responses for Each Logical Case in Each Frequency Group (Exp.7).

	Infrequent Group				*Frequent* Group			
Inference Type	MP	DA	AC	MT	MP	DA	AC	MT
If P then Q	<u>91</u>	63	<u>81</u>	69	<u>100</u>	69	<u>84</u>	59
If P then not Q	<u>97</u>	38	56	<u>75</u>	<u>97</u>	53	66	<u>75</u>
If not P then Q	72	<u>63</u>	<u>94</u>	53	63	<u>78</u>	<u>91</u>	41
If not P then not Q	56	<u>28</u>	41	<u>50</u>	53	<u>50</u>	63	<u>47</u>

NB matching cases underlined.

API

A true measure of matching bias is observed by calculating the API. It will be remembered that affirmative premises are ones where the categorical premise matches the referred clause. The index as before, is calculated by subtracting the amount of endorsements of inferences with mismatching (or negative categorical premises) from the amount of those with matching (or affirmative premises). A one-sample *t*-test is calculated on the indices for each group to see if it is significantly different from zero (an indicator of the presence of matching bias). A two-sample t-test is then used to see if there are significant matching differences between groups (*infrequent* and *frequent* words). The results are shown in Table 6.6.

Table 6.6. T-test Values for the API Scores in Each Frequency Group (Exp.7).

	t Value
Infrequent Group [a]	3.09 *
Frequent Group [b]	3.59 *
Between Groups	-0.41
	(*df*=61)

* Denotes significance
[a] *one-tailed* test for significance (P<.01)
[b] *two-tailed* test for significance (P<.01) between conditions

The t-test results show that there was a significant amount of matching bias taking place in each of the two groups but there was no significant difference in the amount of matching bias between groups. Unlike the previous API results from Experiments 4 and 6, it would appear that there are similar levels of matching bias occurring in both groups (using both frequent and infrequent words). The effects using large and small categories were not resembled in the present study, suggesting that previous findings with concept sizes can not be accounted for just by familiarity or relatedness measures.

NCI

The NCI for each condition were calculated in order to see if there was a significant amount of negative conclusion bias occurring in both groups (see Table 6.7). Unlike the findings from Experiment 4 and 6, there is a significant amount of negative conclusion bias occurring in both groups. The present findings concerning the NCI reflect those found in

Experiment 5, where it was deemed that the category size effect was ineffective. In that experiment, all NCI values reached significance. The present findings support the contention that differences in NCI scores found previously between conditions occurred because of the concept size manipulation and not because of extraneous influences concerning word relatedness.

Table 6.7. T-test Values for the NCI Scores in Each Frequency Group (Exp.7).

	T value
Infrequent Group	2.68 *
Frequent Group	2.60 *
Between Groups	0.14
	(*df*=61)

* Denotes significance
[a] *one-tailed* test for significance (P<.01)
[b] *two-tailed* test for significance (P<.01) between conditions

Discussion

Latencies & Responses

The proportion of response *latencies* seems to inversely correlate with proportion of responses, as was the case in the three previous conditional inference experiments (excluding Exp. 6, *indeterminate inferences*). It would seem that reasoners are more likely to respond quickly if they deem the appropriate response to be one of endorsement ('follows'). This is especially true of explicit affirmation problems, which elicited the most 'follows' responses and the least amount of response time. This effect seems more apparent in the present experiment (and Experiments 4 and 5) than it was in Experiment 6, where more control was in place for the overall problem structures. In Experiment 6, the categorical premises for all four inference-types included explicit and implicit negation, thus controlling for the presence or absence of negation throughout the whole problem structure. This resulted in a more equal amount of inference endorsements and response latencies. It might therefore be presumed that normally large proportions of responses for explicit affirmation problems might be partly due to the actual problem structure itself.

Matching Bias

The results showed the rather typical and robust findings in terms of an overall matching bias effect. It is apparent that in this study, and in the previous Experiments 4 and 5, the largest difference in latencies and response rates between the four inference types occurs with explicit affirmation problems. Latencies and response rates to these problems appear to be mainly accountable for the significant main effects of 'matching bias' in terms of referencing and inference. The three other inference types (explicit denial, implicit affirmation, and implicit denial) show a far lesser degree of response differences.

Matching Relationship

The experiment aimed to clarify the role of the relationship of the categorical premise/referred clause pairing in terms of matching. It was previously seen that the concept size manipulation elicited differences in response rates between conditions with regard to inferences whose categorical premise/referred clause concept pairings matched. In light of previous memory research, it was unclear whether these differences occurred because of the intended concept size manipulation, or whether extraneous variables such as word relatedness could account for the effects. If these extraneous variables were to account for the effects, then it would pose serious problems for the mental activation model. This was not found to be the case.

Word Relatedness

Relatedness (or typicality) of the concept pairings forming a matching inference does not affect the amount of matching that takes place. It would seem that when the concept sizes remain constant (superordinate category and an instance of that category), word relatedness (frequency) had little or no effect on the amount of matching or mismatching endorsements that are given. For instance, in both conditions, reasoners were given inferences where the referred clause was a superordinate category of say 'an animal'. In the low frequency group the categorical premise would have been 'reindeer' whereas in the high frequency group it would have been 'dog'. It appears that both low and high frequency concepts match the superordinate concept 'animal' to the same extent. Whereas in Experiment 6, the superordinate category in the *m-l* group was 'animal' paired with 'reptile' and in the *s-l* group the pairing would have been 'animal' and 'adder'. This difference in the overall concept size of the two items produced a difference in the amount of matching. Whilst it may have been argued that 'adder' was a term more related to 'animal' than was 'reptile', the present findings suggest that the effect was more likely to have been caused by the original concept size manipulation rather than any effects of typicality or relatedness.

Summary

These final two experiments have clarified the initial inference task findings found in Experiments 4 and 5. Experiment 6 has replicated the concept size effects that were previously found in Experiment 4 but were not replicated in Experiment 5. By controlling for the relative size of the concepts between conditions has again produced the concept size effect on matching bias. The use of indeterminate problems in Experiment 6 has also given an indication that matching concepts in a problem may cause a facilitatory effect of processing the problem which was also proposed in Experiment 4 for the sentence verification task. It would seem that matching bias is not always indicated by a larger frequency of endorsements. The findings from Experiment 6 have also raised issues concerning the problem-structure differences using the standard negations paradigm. By making comparisons between inferences that control for the presence or absence of negations in the major premise and conclusion, there is found to be a more even amount in the frequency of endorsements for the four types of inference (explicit/implicit affirmation/denial). That is, there is not such a greater frequency of explicit affirmation inference endorsements compared with the other inference types.

Experiment 7 has resolved a possible criticism with regard to a confounding effect of word frequency rather than concept size. It was apparent that the word frequency manipulation does not affect matching bias in the way that the concept size manipulation does. This supports the contention that concept size directly affects the spread of mental activation needed to represent the problem. The amount of activation power available to memory can still be put forward as an explanation of possible relevance effects associated with the matching bias phenomenon in line with the mental activation model proposed in Chapter 3. It would seem that mere comprehension or familiarity of the materials is an insufficient explanation of the observed concept size effects and has thus offered more support for the mental activation model.

Chapter 7

Discussion

It would be useful before beginning the discussion to briefly outline the major findings of both the selection task and conditional inference task studies.

The findings from the selection task studies showed:

- Matching bias occurs for both lexical and semantic concepts.
- A likelihood rating scale may be a better indicator of card relevance than the frequency of card selections (they do not correlate).
- When all the cards are semantically similar, none of them are seen as significantly likely or unlikely to falsify the rule.

The findings from the conditional inference tasks showed:

- The matching bias effect disappears when using large concepts.
- The sentence verification task shows matching bias.
- Matching bias is task specific and does not always show as a tendency to *endorse* the inference but can also show as a tendency to *reject* the inference.

Matching Bias

In both the selection task and conditional inference experiments, it has been consistently seen that the matching bias phenomenon still occurs even when items do not lexically match. This had not been specifically shown until now. In fact, there has been some recent confusion as to what constitutes the matching relationship between concepts, as was pointed out in Chapter 3 with specific reference to Yama (2001). The uncertainty as to what constitutes matching is highlighted in Yama's first selection task experiment, where the consequent clause stated that a card could have a number from '1-1000' printed on it and the number '376' was used for the *matching* consequent card. Yama (2001, Pg.304) concluded that, "...it is doubtful if a number such as 376 exactly matched or not the statement '1-1000', even though participants can easily recognise that the number is an element of the set". He then went on to run a third experiment using a number on the cards that *lexically* matched ('2000') the set mentioned in the consequent clause ('11-2000'). The results from the present study conclusively show that if an item is seen as *representing* an instance from a category (or set), then this alone is sufficient for the matching bias phenomenon to take place. Items do not have to lexically match.

The issue as to what constitutes a matching relationship has been overlooked for some thirty years – ever since the matching bias phenomenon was discovered. Because of this, current reasoning theories have not incorporated any explanation or predictions concerning the possible manipulation of the size of the concepts in a problem. They have merely assumed that matching concepts would be lexically identical and therefore the issue of differing concept sizes has not been pertinent. A brief overview of the three main theories that address matching bias will be given in order to recap on the proposed mechanisms they describe for the phenomenon.

Representational Accounts of Matching Bias

Heuristic-Analytic & Dual Process Accounts

The basic tenet of Evans' (1984,1989) 'heuristic-analytic' approach and Evans' (2002) 'dual process' account describes how a problem has to be represented before any conscious hypothetical thinking about the problem can occur. Thus, there is seen to be two stages in the thought processes about a problem (Evans adopted Stanovich's, 1999, System1 and System2 terminology). Systematic biases such as matching are seen to derive from selective representation processes at the unconscious System1 level. If the problem is incorrectly modelled or represented then it is unlikely that correct responding to the problem can be achieved. Formal, conscious System2 thought processes can only operate on the initial representation of the problem created by the unconscious System1 process.

Therefore matching bias is seen as systematic selection errors at the System1 level caused by unconscious perceptions of what is seen to be relevant aspects of the problem.

The present line of studies has focussed on the System1 or representational stage of a problem. The general question was posed as to what is *relevance* in relation to the problem materials? Although Evans (1984,1989) proposed that items in a problem are seen as more relevant if they match the concepts named in the premise, no account of the mechanisms that cause a heightened perception of relevance have been specified. It is accepted that reasoners are not consciously aware of this bias as otherwise it could be presumed that they would rectify this seemingly clumsy response to a logical problem. The question arises as to what are these unconscious processes that pick out relevant aspects of a problem? It has been argued that any answer to this question must fall within the realm of unconscious human perceptual processes that lead to the comprehension and representation of the problem.

Mental Model Theory

The mental model theory of Johnson-Laird and Byrne (1991) proposed that an initial model of the premise is constructed against which tentative conclusions are formed. Attempts are made at formulating alternative models in which the tentative conclusion does not hold. If no such model can be constructed, the conclusion is seen as valid. One of the fundamental aspects of the Johnson-Laird and Byrne (1991) mental model theory was that all the concepts are explicitly represented in the initial construction of the model. For instance, if a negated concept appears in the premise (e.g. if A then NOT 2) the model would represent the assertion 'not 2' and its corresponding affirmative proposition '2'.

Johnson-Laird and Byrne (2002) have since amended the model so that only the true assertion 'not 2' is represented in the model because the original model had difficulties explaining the *double-negation* effect (Evans & Handley, 1999). The matching bias phenomenon is now seen as a 'mismatching' effect caused by the use of implicit negations, in that reasoners are seen to have difficulty equating that a '3' for instance, refutes a '2' or that a '3' affirms 'not 2'. Johnson-Laird and Byrne suggest that working memory is fundamentally necessary in creating the different models and that the more models that can be created from the given premise, the more the problem will be perceived as difficult. Problem difficulty and biased responses are seen as stemming from working memory limitations and the mechanisms that form the initial representation of the problem. Again, they do not specify the perceptual mechanisms that underlie the formation of the initial explicit model of the problem.

Evans (1991,1993) has suggested that this initial explicit representational process in the mental model theory coincides with the System1 account in that it is an unconscious and perhaps perceptual process that creates the model. Again, as in the dual-process account, Johnson-Laird and Byrne have not stipulated in their theory what the mechanisms are for these unconscious processes and how they seem to be biased towards selecting matching aspects of the problem. It was hypothesised in Chapter 3 that the process of the mental activation of the concepts in memory could offer an explanation for matching responses in a conditional reasoning problem. Based on theories of a distributed memory and how these theories offer an explanation for priming phenomenon, it was suggested that similar unconscious processes might also account for matching bias. It was suggested that heightened perceptions of relevance of certain problem items over others might be caused by the residual trace activations of those concepts in memory. Thus the present line of research was intended to offer an explanation for the underlying mechanisms that cause unconscious perceptions of relevance rather than as a test of the two theories outlined above. As the two theories offer no detailed account for these mechanisms, they therefore do not put forward any refutable predictions concerning the concept size manipulations that took place in the present study.

Probabilities of the Concepts

Optimal Data Selection

The only other current theory of matching bias in conditional reasoning problems is offered by the rational-analysis view of Oaksford and Chater (1994). Instead of suggesting representational factors as causing matching bias in terms of how the problem is perceived, this account offers an alternative explanation in terms of the probabilities represented by the problem concepts. For the selection task, they suggest that matching cards represent rare information and are therefore more likely to be chosen. The assumption is that the non-matching cards represent negated contrast sets i.e. 'not a cat' represents a much larger set than the actual instance 'cat'. Therefore, the probability that something is a cat is much lower than the probability that something is not a cat. Rare information is seen as more informative and is therefore more likely to be chosen.

Oaksford, Chater and Larkin (2000) offer a similar, but more complex account to explain polarity biases in the conditional inference task. This account explains patterns of conditional inference endorsement in the respect that reasoners are seen to endorse more inferences with affirmative categorical premises and more inferences with negative conclusions – depending on the probability of the conclusion given the categorical premise. With implicitly negated inferences, they claim the implicit negation is not re-coded into its contrast set. Instead, the implicitly negated clause is seen as representing a lower probability of occurrence than an explicitly negated clause. Oaksford (2002) gives the following example: Given the two inferences, 'if A then 2, *K*, therefore not 2' and 'if A then 2, *not A*, therefore not 2', the probabilities of the conclusion given the categorical premise differ between the two inferences. Such that, P(not 2|K) is less than P(not 2|not A).

This is offered as an explanation for the differences in the frequency of endorsements between implicitly negated inferences and explicitly negated ones. It would thus be expected that an implicitly negated concept that was large rather than small would increase the probability of the inference being endorsed. The mental activation model predicts the converse in that inferences made up of large concepts would be less likely to be endorsed.

Instead of seeking rare and informative information, reasoners are seen as endorsing high probability conclusions. It was pointed out previously in the experiment chapters that affirmative premise bias is equivalent to a measure of matching bias, especially when using implicitly negated minor premises. By predicting affirmative premise bias they are thus predicting the matching bias effect in conditional inference tasks but unlike the mental activation account, they predict more matching bias with inferences made up of large, and hence high probability, concepts.

The present study aimed to delineate the opposing ideas that; on the one hand, biases are caused by some underlying mechanism causing unequal perceptions of relevance between the concepts used in conditional inference tasks; and on the other hand, biases are caused by some formal or perhaps intuitive calculations of the probability of occurrence of the concepts. Some of the predictions of the proposed mental activation model paralleled those of the optimal data selection account. For instance, the mental activation model proposed that large concepts would reduce matching bias in that their mental activation would be spread to many other connected concepts. This spread of activation would reduce the *activation level* of that concept and therefore reduce its perceived relevance. Using the ODS account it could be argued that larger concepts increase the probability of that item occurring in the real world e.g. something is more likely to be a *plant* than it is to be an *oak*. Reasoners for instance in the selection task, would be less likely to make a matching response with 'plant' as it is seen as less rare and informative.

Also, the two accounts make similar predictions with regard to the use of implicit negations. The mental activation model predicts that implicit negations will not be seen as relevant as the concept that represents an implicit negation does not overlap the activation of the concept mentioned in the premise. The ODS account points out that the implicit negation represents the contrast set and therefore, because this set represents a higher probability of occurrence it is seen as less informative than information that is rare. This assumption lies entirely on the notion that reasoners re-code an implicit negation (e.g. 'dog') into an explicit one (e.g. 'not a cat') in the context of the problem – otherwise it wouldn't represent such a large contrasting set. Oaksford (2002) has suggested that the use of explicit negations reduces matching bias through a simple *matching₂* strategy whereby the whole clause on the card e.g. 'not 2', is matched to the premise clause 'not 2'. Therefore, the *matching₁* '2' card is less likely to be chosen. It is not clear from these accounts if processing an implicit negation is more or less difficult than processing an explicit negation.

The approach of the present study was mainly to explore, or at least better specify, possible underlying mechanisms that could explain relevance effects. Whilst similar predictions were also made in line with the ODS account, the materials used could also show that the two accounts can be distinguished. The main findings will now be addressed in relation to the current theories mentioned above and it will be shown that the results offer support for the mental activation model in that it makes surprising predictions that are not currently addressed in the present reasoning literature. The results have also transpired interesting questions with regard to the present status of the field.

The discussion of the experimental findings will be addressed in two sections. The first section will look at *within* task differences in response rates. It will be remembered that participants generally received 16 inferences in line with the negations paradigm (see Chapter 3, Table 3.1). All these inferences differ in respect to the placement of negations and whether the minor premise constituted the antecedent or consequent item (forward or backward inferences). Schroyens, Schaeken and d'Ydewalle (2001) placed the 16 inferences into four categories: explicit affirmation, explicit denial, implicit affirmation and implicit denial. The analysis of responses concerning matching bias and negation effects usually takes place with an analysis of these inference structures within the task. For instance, matching bias is seen when there are more endorsements to explicit rather than implicit type inferences.

The second section will address what is being called here *between* task variances in response rates. The between task variable of interest is primarily the concept size manipulation. The concept size analyses compare like-with-like in respect to the actual inference structures (i.e. placement of negations, logic etc.). This section will therefore look at categorical and semantic relatedness effects and some of the broader issues that have arisen with manipulating concept sizes. It is only in this section that the mental activation model can be delineated from current theories. Up until this point, the mental activation model predictions coincide with current theoretical predictions and these will be outlined in the first section. It is only with experimental support for surprising predictions can the mental activation model be seen as a true alternative explanation for response biases seen in conditional reasoning tasks. These predictions stem from the use of the categorical/semantic relatedness effects and the subsequent concept size manipulation.

Within Problem Aspects of Matching, Explicit/Implicit Referencing & Negation

Matching

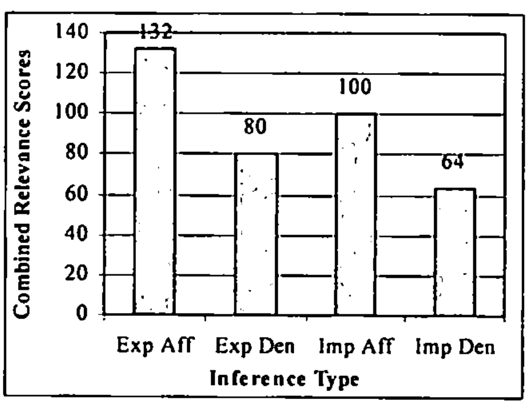
The results from all the experiments showed consistent matching bias trends using problem materials that did not lexically match. Matching bias is seen in the selection task when cards that match the items named in the premise are chosen more often than cards that do not match, irrespective of the logical requirement designated by the problem. With the conditional inference task, matching can be seen when there are more endorsements to inferences whose minor premise item matches the respective item in the major premise. The matching effect is solely dependent on the use of implicit negations – which were used throughout this study (with the exception of Experiment 6). For instance, given the premise, 'If there is an A one the left side then there is a 2 on the right side', an explicitly matching antecedent case would be 'A' whereas a mismatching one would be 'B'. This led Schroyens et al (2000) to label the effect as '*mismatching*' bias. According to Schroyens et al, Evans (1998) and Johnson-Laird and Byrne (2002), implicit categorical premises (e.g. 'K') are not seen as being relevant to the referred clause, with or without negation (e.g. 'A' or 'not A').

The mental activation model outlined in Chapter 3 proposed that items do not have to match (or mismatch) as such in order to create the biased responses (It could be argued that essentially all the items used in this study have 'mismatched' – at least lexically). The activation account suggested that as long as some items in a problem overlapped in their activation of a semantic network, they would appear more relevant than items that did not overlap in activation. It was argued that trace memory activations caused 'incremental' amounts of activation to occur so that if the same semantic network or concepts were activated more than once, the level of activation would be higher than that occurring from only one activation. Matching cases overlap – mismatching ones do not.

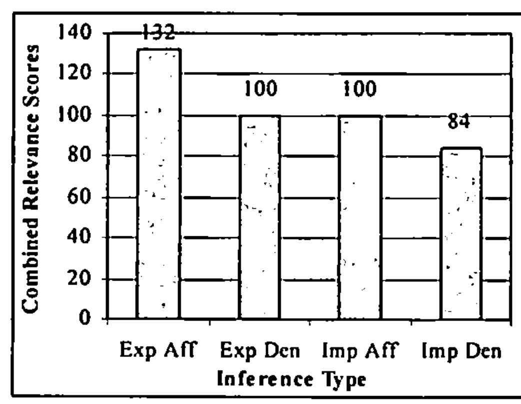
On the surface, it could be seen that the mental activation account is no more informative than existing theories. But the mental activation model makes more precise predictions concerning the standard pattern of results that were reflected in the present study. The following figures from Chapter 3 have been reproduced here for ease of reference.

Figures 3.1 & 3.2 Accumulated Relevance Scores for Explicit/Implicit Affirmation/Denial Inferences (Reproduced from Chapter 3).

Baseline Model



Reactivation Model



The mental activation model appointed a simple relevance-scoring scheme to each of the 16 negations-paradigm conditional inferences. Relevance scores were attributed by the size of each concept and whether the items in the categorical premise and conclusion overlapped in their activation with the activation of the concepts mentioned in the major premise. These 'raw' relevance scores were accumulated for each of the four inference types and can be seen in the first figure. It must be emphasised that the first figure is based solely on the structure and content of the inference. It does not take into account any formal System2 processing.

The second figure shows essentially the raw relevance scores but without including the score for the activation of the concept that represents the conclusion. This was felt to represent participants' inhibition of the given conclusion in order to try to formulate their own conclusion. These accumulated relevance scores therefore include a primitive allowance for partial System2 reasoning processes and by doing so the model also caters for apparent double-negation effects. Based solely on a rather crude but consistent scoring method, the figures remarkably resemble the actual patterns of responses that are generally found using the conditional inference task. The principle of relevance that is modelled by the mental activation approach not only reflects the experimental findings, but also satisfies the principle for simplicity of explanation. For instance, the first figure is almost an exact duplicate of the findings from Experiment 4 (Chapter 5, Figure 5.6).

All the theories of matching bias outlined previously that argue for relevance effects (explicit/implicit negation) do not, for instance explain why more responses are obtained from explicit affirmation problems. In fact, they do not predict this pattern of responses. They merely predict that there will be more matching (explicit) than mismatching (implicit) responses and therefore implying there should be no difference in the amounts of endorsements to both explicit problem types, and likewise for both implicit problem types. Such a pattern of responses is seldom if at all found.

The rational analysis view of Oaksford and Chater (1994) has initial difficulties with the findings concerning explicit/implicit negation. In the selection task for instance, they suggest that mismatching cards represent a contrast set of instances. Using the example, 'If there is a tree on the left side there is a cat on the right side', with the two consequent cards 'cat' and 'dog' they suggest the 'dog' represents the contrast set 'not a cat'. This contrast set is seen as representing a larger probability of occurrence in that something is more likely to be 'not a cat' than it is to be 'a cat'. This they say explains matching bias in that reasoners will prefer rare information as it is more informative. But what happens if a card states 'not a cat' rather than 'dog'? Previous findings show that the use of such explicit negations eradicates the matching bias effect.

Matching$_1$ and Matching$_2$

Oaksford (2002) believes the matching bias effect disappears using explicit negations because of a lexical matching *strategy*. He proposes that the standard matching phenomenon using implicit negations be classified as *matching$_1$*. To account for the lack of a matching bias effect using explicit negations in both the truth table and selection tasks, Oaksford proposes that a *matching$_2$* strategy may be taking place. Here, matching is seen to take place between the whole clause i.e. 'not p' matches 'not p'. Hence, using explicit negations in a task causes all the items to appear relevant. Oaksford states that this explanation is only pertinent to the truth table task and selection task but not the conditional inference task. In the selection task for instance participants re-code the implicitly negated card into the contrast set (i.e. 'dog' = 'not a cat'). In the conditional inference task he suggests reasoners no longer do this because the task is less ambiguous.

It is not parsimonious to have two separate accounts of the findings concerning implicit and explicit negations. On the one hand you have contrast-sets and probabilities (matching$_1$) and on the other hand you have matching$_2$. Why would reasoners adopt two different strategies? If they *match* using explicit negations then they are likely to match using implicit negations. The current selection task findings (Exps. 1,2,& 3) do not support this contention in that it has been seen that items do not have to be lexically identical to invoke matching bias. It is not clear why reasoners would adopt a lexical strategy of matching$_2$ whole clauses only when explicit negations are used. It is clear from the present results that the *concepts* mentioned in the problem (irrespective of negation) cause matching bias even if they do not lexically match. Matching thus goes far beyond simple lexical matching strategies. But what Oaksford is actually suggesting is very similar to the mental activation account. For instance in the selection task using explicit negations, Oaksford claims that given the premise 'if not p then q' the 'not p' and 'q' cards *completely match* and the 'p' and 'not q' cards *completely mismatch*. The pairings of 'p,q' and 'not p, not q' are classed as *partial* matches. The matching of both cards is taken into account to give a three-point scale of the degree of matching taking place with the inference.

It will be remembered from Chapter 3 that the mental activation model is based on the same principle. Instead of using the term 'match', degrees of mental activation overlap were instead considered but which essentially add up to the same thing. It was pointed out that the concepts could *totally* overlap, *partially* overlap or there could be *no* overlap. Oaksford therefore is essentially agreeing to the principle of the mental activation account for explicit negations. The mental activation model does not have to be altered or amended to also account for the matching bias findings using implicit negations whereas Oaksford suggests a completely different mechanism for these matching$_I$ effects. The mental activation model also predicts the *conditional inference* findings without any alteration whereas Oaksford again suggests a further mechanism concerning the probability of the conclusion given the categorical premise.

Negative Conclusion Bias and Double Negation

Other mechanisms such as double-negation effects are offered in addition to relevance effects in order to explain this normal pattern of findings. Evans and Handley's (1999) *double hurdle theory* states that the first part of reasoning concerns perceptions of relevant aspects of the task. If reasoners deem the wrong aspects of the task (i.e. matching ones) to be relevant then they are likely to fall at this hurdle. If they get passed this hurdle, then they may face difficulty at the second hurdle in denying an explicit negative (e.g. seeing that not 'not a cat' is actually 'a cat'). Unconscious System1 processing causes the first representational stage whereas the latter second-hurdle stage in trying to formally calculate a conclusion to the inference is undertaken by System2 processes. The two figures above, parallel this two-stage principle but in fundamentally different ways. Figure 1 shows a pattern of responses that are solely determined by the perceived relevance of the concepts mentioned in the problem. The second figure on the other hand incorporates a similar System2 account but not concerning difficulty with negations, but simply an inhibitory activation of the conclusion. This inhibition must come about through conscious System2 attempts by reasoners at trying to generate their own conclusion. The necessity for such an operation would surely be to inhibit activation of the concept used for the given conclusion. So, whilst this inhibitory process is also seen as a System2 process, the *apparent* double-negation effect can appear to stem solely through inhibition of the conclusion and not attempts at trying to calculate double-negatives.

The two figures above show how response patterns can change from one experiment to the other simply by the extraneous variable of whether participants actually try to generate their own conclusions and solve the problem. The mental activation model predicts experimental response patterns lying between those seen in Figures 3.1 and 3.2. Depending on whether conscious System2 processes are adopted by participants in trying to solve the problem, will the pattern of responses swing from the baseline model (System1) to the reactivation model (System2). The mental activation model predicts that double-negation effects will be weak when System2 conscious thought processes are not applied to the task and that the baseline response (no formal cognitive effort) to a problem is dependent on System1 relevance perceptions. The response patterns from the conditional inference tasks in Experiments 3, 4 and 6 all fall within these two predicted patterns outlined in the above two figures. The mental activation model accurately predicts response proportions *solely* by reference to perceptions of relevance caused by overlapping activations of the problem concepts. It therefore offers a far more simplistic and detailed account of *all* the general findings concerning explicit/implicit referencing and negation effects in abstract tasks.

Evans and Handley (1999) used implicit negations for the categorical premise in a conditional inference task. They found that responses to inferences with implicitly negated premises were suppressed. This is essentially the matching (or mismatching) bias phenomenon. These findings were in contrast to those found by Evans, Clibbens and Rood (1995) who used conditional inferences with explicit negations for the categorical premises. Using explicit negations, Evans et al found no evidence of affirmative premise bias but did find evidence of negative conclusion bias but primarily on denial inferences which required the calculation of a 'double negation'. They concluded that the bias was not a tendency to endorse more inferences that had negative conclusions, but a *suppression* of responses to those inferences that required this step of double negation (denial inferences with affirmative conclusions).

The current line of studies using the conditional inference task has shown clear evidence of a double-negation effect. In all four inference task experiments it is seen that there are more responses to denial inferences that do not require the calculation of a double negative. In all experiments the proportions of responses follow the following patterns; DA1+DA3+MT1+MT2>DA2+DA4+MT3+MT4.

The memory activation model predicts the double-negation effect for different reasons to Evans and Handley. The mental activation model predicts the double negation effect by an *increase* in the frequency of endorsements of denial inferences with 'negative' conclusions (see Table 3.4, Chapter 3). Evans and Handley predict that the effect is caused by a *decrease* in the frequency of endorsements of denial inferences with 'affirmative' conclusions. The explanation stated for this in the mental activation model was a possible attempt by reasoners to generate their own conclusions. To do this, it was proposed that reasoners inhibit activation of the concept mentioned in the given conclusion and concentrate harder on the categorical premise concept (thereby increasing its activation level). By modelling this approach in terms of *relevance scores* the mental activation model encapsulates the double-negation effect.

With regard to the overall NCI values reported in the present study they would seem to support the findings of Evans et al (1995) and Evans and Handley (1999) in that when there is a significant API measure there appears also to be a non-significant NCI measure. (The discussion of the concept size manipulation and the effect on the NCI will be left until the next section). Evans, Clibbens and Rood (1995) found that by using explicit negations in the conditional inference task there was found to be significant NCI but weak API scores. When Evans and Handley introduced *implicit* negations into the task, generally the opposite was found in that there were significant API's but non-significant NCI's. The reason for a lack of occurrence of both significant API and NCI scores may be an artefact in that the two measures are not entirely independent. For instance, both indices contain the measure of taking the amount of endorsements to inferences with both *negative premises* and *affirmative conclusions* away from endorsements to inferences with both *affirmative premises* and *negative conclusions*. A strong leaning towards affirmative premise bias would reduce the measure of negative conclusion bias and vice-versa.

As previously stated, Evans and Handley believe the negative conclusion effect to be mainly caused by double-negation difficulties and is therefore more pronounced for denial rather than affirmation inferences. This is supported by Schroyens, Schaeken and d'Ydewalle (2001) in their meta-analysis of several conditional reasoning experiments. Their results showed that reasoners make consistently fewer endorsements to denial inferences that have affirmative conclusions (negative inferential-clauses) than those that have negative conclusions. Their results are indicative of a reliable double-negation effect.

The Matching Relationship

It has been seen in all the seven experiments in the present study that matching bias occurs using concepts that do not lexically match. It was pointed out at the start of this chapter that there had been a lack of clarity as to what matching actually meant. So far, the present study has indicated that concepts only have to be semantically related in order to match. But this poses a further question as to what *semantically related* actually means? On the surface it seems like a reasonably descriptive label for the named concepts in the problems that have been developed in the current study, but it introduces concerns when looking at the whole *clause* in a problem especially with regard to negation. For instance, why would the clauses 'cat' and 'not a cat' be semantically related (matching) whereas 'dog' and 'not a cat' (mismatch) would not? Surely 'dog' and 'not a cat' are semantically related in the context of a problem because they share the same logical meaning, although the relationship between them is one-way in that 'a dog' is 'not a cat' but 'not a cat' is not necessarily 'a dog'. The only reasonable answer to this paradox is to assume that the meanings or alternative concepts associated with explicit negations are not fully represented.

This explanation to some extent already exists in the literature whereby Evans (1998) has claimed that the normal linguistic processes associated with negations direct attention to the negated constituent. Therefore, in the case of 'not a cat', attention is drawn to 'cat' and not the almost infinite amount of concepts the clause actually refers to. Likewise, Johnson-Laird and Byrne (2002) have suggested the mental model invoked by the problem only explicitly represents negated instances but flagged with a negation marker. Only Oaksford (2002) argues that an explicit negation directs attention to the psychologically plausible contrast-class set. Thus, 'not a cat' might direct attention to 'dog' in which case reasoners would be likely to see the connection.

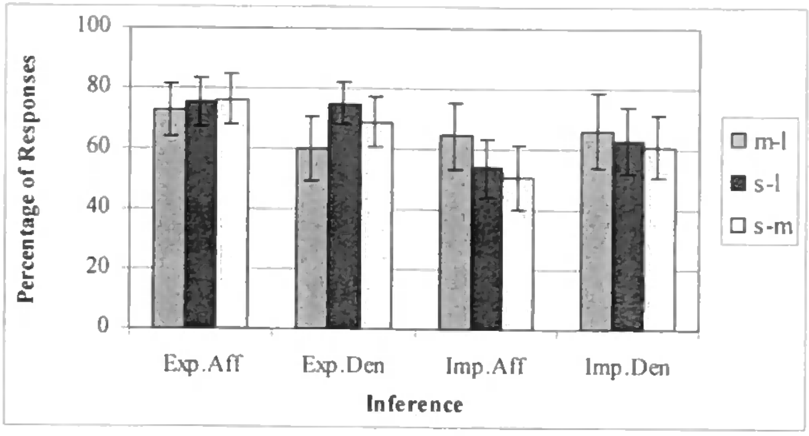
The mental activation model describes a third line of thought. That the word 'not' causes an *inhibited* activation of the concept that follows it and evidence for this can be found in studies such as MacDonald and Just (1989) and Lea and Mulligan (2002)(see Chapter 3, *Relevance & Negation*). By allowing for the idea of a suppressed activation of the negated concept, the idea of a *partial* match can be introduced. By doing this, the standard pattern of findings seen in Experiment 3 (Fig. 5.6), Experiment 4 (Fig. 5.10) and Experiment 6 (Fig. 6.2) can be totally reflected by the relevance scores shown in the two figures above (Figs. 3.1 & 3.2). So by pointing out that the concepts 'cat' and 'cat' *totally* match (overlap), 'cat' and 'not a cat' *partially* match (overlap) and 'dog' and 'cat/not a cat' *do not* match (overlap), the actual proportions of responses found in the experiments can be wholly accounted for.

This cannot be achieved if items are only considered to match or not i.e. 'cat' equivalently matches 'cat/not a cat'. If there were only matching or mismatching cases then there should be equivalent amounts of endorsements to explicit affirmation/denial type problems and equivalent amounts of endorsements to implicit affirmation/denial type problems. This has not been found in the present study or elsewhere and so complex alternative explanations have to also be offered by current reasoning theories in *addition* to explanations for matching.

Problem Structure

The apparent exception to the pattern of responses predicted for the conditional inference task is seen in Experiment 6 (Chapter 6, Fig. 6.4).

Figure 6.4. Percentage of Responses for Determinate Inferences for All Three Category Pairings. (Reproduced from Chapter 6)



The patterns of responses do not mirror those predicted in the above two figures (3.1 & 3.2), in that there was found to be a more uniform frequency of responses for the four inference types. There is a clear reason for this. The task did not utilise all 16 negations-paradigm inferences – only half of them. This experiment utilised the inferences which had previously been called 'implicit' and used the same inferences to create explicit ones. therefore, the explicit/implicit distinction was made solely on the same inference structures. For instance, an explicit modus-ponens inference in the experiment would have read:

If there is NOT an adder on the left then there is a diamond on the right.

There is NOT a reptile on the left.

Therefore there is a diamond on the right.

For the implicit inference of equal problem structure, the minor premise would have read:

There is a flower on the left

Normally, the explicit/implicit negation differences (matching bias) have been observed between inferences with different problem structures in that the major premise polarities have also been different i.e. comparing 'If A then 2. A therefore 2' with 'If NOT A then 2. B therefore 2'. It can be seen from Experiment 6 that when the major premise structure is maintained together with the conclusion, a more even amount of responses between the four inference types is achieved. This indicates that *partial* matching must be taking place. In the normal paradigm using the 16 inferences with implicit negation there is found to be a predominately large amount of explicit affirmation endorsements (Exps 4,5 & 7). It is argued here that this occurs because all the items that match using these types of explicit affirmation problems are *affirmative* and therefore *totally* match i.e. 'A with A' and '2 with 2' (see Chapter 3, Table 3.1). None of the categorical/referred clause pairings in Experiment 6 totally match and therefore do not produce disproportionately large amounts of explicit affirmation endorsements. This shows further support for the mental activation account of partial matching.

What is Relevance?

Two matters arise concerning what is relevance. Are selection task choices an indicator that the card is relevant to falsifying the rule? And, are *endorsements* to conditional inferences an indicator of relevance/matching bias? It has always been assumed that perceptions of relevance transpire in conditional inference tasks through participants' responses. In the selection task, cards are seen to be perceived by participants as relevant to falsifying the rule, if they are chosen. In the conditional inference task, matching bias (and therefore heightened perceptions of relevance) are thought to be indicated by an endorsement ('follows' response) of the inference that matches and a refutation ('does not follow') of an inference that does not match.

It was seen in Experiment 3, using both card selections and a *likelihood rating scale* that there is a disparity between selecting a card to falsify a rule and rating the likelihood that the card could falsify the rule (see Chapter 4, Figs. 4.4 & 4.5). The usual pattern of card selections was revealed in the experiment in that participants predominately select the matching TA and TC cards. But the likelihood ratings are more pronounced for the matching TA and FC cards. That is, overall participants saw the matching FC card as more likely than the matching TC card to falsify the premise. Yet they selected the matching TC card more often. It would seem that there may be some pragmatic factors with the task that cause participants to embrace a verification strategy when selecting cards, but actually see the card's true relevance to the rule when asked to rate how likely the card could falsify the rule.

The rating scale undoubtedly shows matching bias in a very clear way. Mismatching cards, irrespective of their logical function, are predominately rated as less likely to falsify the rule. As the rating scale also gleans information from the cards that are not chosen, it can be seen as a true indicator of card relevance in the task. By comparison then, a measurement based entirely on the cards that are selected can be considered a flawed measure of the cards' relevance. It was pointed out in Chapter 3 that card selections alone may not be a true indicator of the perceptions of relevance that are taking place within the task. The results from Experiment 3 indicate that further consideration may need to be applied when interpreting the results from selection tasks.

With further regard to the connection between responses and relevance (matching), Experiment 6 showed an unusual effect concerning matching concepts in the inference task. It would appear that matching bias could be observed in the amount of inference *refutations* as well as endorsements. Whilst the experiment showed expected matching bias trends using standard inferences, it revealed a *reduction* of endorsements (more 'does not follow' responses) to 'indeterminate' inferences that matched (see Chapter 6, Fig. 6.5). An inference was classed as indeterminate when there was no logically deductive outcome. For example, an indeterminate matching inference would have been:

If there is NOT an apple on the left then there is an adder on the right.

There is a reptile on the right.

Therefore there is an apple on the left.

Whereas the mismatching equivalent would have had the following minor premise:

There is NOT a sport on the right.

It seems that although all the indeterminate inferences necessitated a 'does not follow' response, participants were more inclined to make this correct response when the items matched. It would seem therefore that matching bias might be a facilitatory effect. If it is accepted that using a biconditional interpretation of the premise, all the standard *determinate* inferences are correct and therefore require a 'follows' response. Matching cases could therefore be seen to facilitate this process. If the correct response should be a refutation of the inference, as is the case with the indeterminate problems, then matching concepts would also seem to facilitate this response.

The issue then arises as to whether matching bias is an indicator of relevance or whether it facilitates the processing of the problem. Matching cards seem to facilitate the participants' ratings of whether a selection task card could falsify the rule and also matching inferences seem to have facilitated the processing of the conditional inference problem. The overlapping activations stated in the mental activation account may be *facilitating* memory and representation of the problem and therefore it may be fallacious to think of matching concepts as increasing perceptions of relevance. *Relevance* may not therefore be a sufficient term to use for these underlying processes.

Categorical & Semantic Relatedness Effects – 'Between' Problem Structures

The concept size manipulation that has taken place throughout the study will now be looked at. These manipulations came about as a direct test of the mental activation model and were essentially unusual or surprising predictions with regard to the patterns of responses that could be obtained from conditional reasoning tasks. It was felt that no current theory of reasoning in their present forms could predict that any differences in responding would occur due to concept size manipulation. Only the proposed mental activation model suggests that using larger concepts, which spread the finite amount of activation power available to memory, would lead to a reduction in the matching bias effect.

Reduced Matching Bias

Experiments 4 and 6 have shown that using large concept sizes does have an effect on matching bias when the matching concepts do not lexically match. The experiments showed that by using large concepts in a conditional inference problem, the matching bias effect significantly diminishes in line with the prediction, although the effect is weak. Matching is more generally observed by a larger proportion of determinate inferences (endorsements) drawn from the problems. Experiment 4 showed that there was a significant overall difference between the large and small conditions. It appeared that by using small concepts in the construction of the problems, participants were more inclined to endorse the inference (or it could be quite conceivable that using large concepts decrease the amount of endorsements).

On closer inspection of the affirmative premise indices (API) the finding that there was no significant API for the s-l concept size inference in the *m-l/s-l* condition in Experiment 4 almost satisfied the prediction. That is, larger concept sizes would reduce the amount of matching bias than that found using smaller concept sizes. The caveat here is that there was still significant matching bias taking place using the largest of the three sets of concept sizes. It appeared the concept size manipulation was disrupting the matching bias effect, but in an unexplained way. This possibly could have been caused by carry over effects of the materials due to the mixed design. Memory manipulation and especially trace activations in a semantic network must be treated cautiously in any experimental design. As attempts at manipulating trace activations and memory were the purpose of the experiment, it could be argued that the design was flawed in that all the matching concepts for each level (in each condition) were activating the same semantic network. Therefore, there is reason to presume that having contact with all three nested-triple concepts (e.g. plant, tree and oak) in one experiment was bound to interfere with the concept size manipulation in terms of trace activation of concepts. It could be that having encountered all three terms, participants trace memories were somehow interfering so that encountering the concept of 'tree' could also be reactivating 'plant' from a previous encounter.

The results from Experiment 6 shed more light on the situation in that there was no significant matching bias (API) on inferences that were constructed using larger concepts. The totally between design controlled for any trace memory anomalies and provided further support that concept size interferes with the matching bias effect. This finding is in contrast to Evans and Handley (1999) who found that matching bias occurred consistently in the conditional inference paradigm using implicit negations. So much so, that they believed matching bias was actually caused by implicitly negated concepts (and hence mismatching cases) not being perceived as relevant to the task. Experiments 4 and 6 show that even when implicit negations are used, the matching bias effect is not always certain.

But rather than show a sporadic or weak matching bias effect, the present study has shown a consistent trend in that matching bias, on problems that are made up of large concepts, diminishes. There is currently no explanation other than the mental activation model that can account for these findings. When incorporating concept size (or area of activation) into the mental activation model the results can be predicted. By the simple assumption that a large concept will theoretically equal the same activation area as a negated smaller concept, and that a negated large concept will have the greatest activation area of all, three sizes of activation area can be incorporated into the model (i.e. small, medium and large). It is not the contention here that a negated smaller concept *actually* equals the activation area of a large concept but that the areas of activation are conceptually relative to each other in this way. By using a very simple algorithm (see Chapter 3), both the general response proportions to explicit/implicit affirmation/denial problems can be explained together with this new finding concerning concept size.

The Relevance of Mismatching Cards

The mental activation account is also supported by the selection task results. It was previously pointed out that a scale measuring participants' likelihood ratings was a better indicator of the perception of all the cards' relevance in the selection task. When this was implemented, there was found to be significant likelihood ratings in the *dissimilar* condition where mismatching cards were not semantically related to any of the concepts mentioned in the premise, but not in the *similar* condition where they were related. This is a new and interesting finding concerning the selection task. In both conditions, the mismatching antecedent and consequent cards did not match their corresponding clauses in the premise. But in the similar condition the mismatching antecedent card did match the consequent clause and vice-versa. Logically this should not have an effect. There was little chance of confusion because the premise clearly distinguished between concepts on the left and right side of the cards and the concepts on the cards were clearly either on the right or left sides. Yet this manipulation made a difference to participants' ratings of each card – even matching ones. The similarity manipulation seemed to interfere with perceptions of *all* the cards. In the similar condition, all the cards were rated as less likely/unlikely to falsify the premise. It appeared that perception of relevance for some cards over others diminished.

This finding was predicted in the memory activation model in that typically in the selection task six concepts are mentioned – two in the premise and four on the cards. Two of the card concepts match the premise items and together with the premise items form two sets of matching pairs. Two of the cards display concepts that do not match any of the other concepts in the problem. It was proposed that the card concepts that form matching pairs with the premise concepts would be perceived as more relevant than the cards that do not match, because of trace memory activations. It was proposed that if a task made the 'mismatching' cards overlap in activation with any of the premise concepts, then perceptions of a matching card appearing more relevant than a mismatching card would diminish.

For instance, in Experiment 3, participants were given the premise, 'if a card has a fruit on the left side, then it does not have a vegetable on the right side'. The cards for both the *similar* and *dissimilar* conditions were thus[3]; 'apple', 'potato (glass)', 'carrot' and 'orange (fox)' where the concepts in parentheses show what was displayed on the cards in the *dissimilar* condition. It was predicted that in the similar condition, both 'apple' and 'orange' overlap in activation of the semantic network 'fruit' and likewise for the two vegetable concepts. So even though the antecedent choice of 'potato' mismatches the concept of 'fruit', its activation will still overlap the consequent concept and it will therefore appear equally as relevant to the problem. This prediction was supported by the likelihood ratings, the results of which cause difficulties for Oaksford and Chater's (1994)

[3] NB The first two concepts were displayed on the left side of the cards with the right side blanked out, whilst the last two concepts were displayed on the right side.

contrast-set account of matching bias on the selection task. Using the above example, they have claimed that both 'potato' and 'glass' would represent the contrast class of 'not a fruit' and therefore would be seen as equally highly probable occurrences. The 'apple' card would therefore be seen as a rare occurrence in *both* the similar and dissimilar conditions.

Admittedly, this account still describes the proportion of card *selections*, but it does not account for the disproportionate likelihood *ratings* using similar and dissimilar terms. Surely, the likelihood ratings are reflective of 'expected information gain' and if the ratings are extreme in either direction – likely or unlikely to falsify the rule then this would be a direct indicator that that particular card is informative. If the overall collective ratings for each card are not significantly different from zero, this must reflect overall uncertainty as to the informativeness of the cards. Why would matching cards be seen as less likely to falsify the rule in the similar condition?

Concept Size Effects

The sentence verification task used in Experiment 4 did not replicate the relatively robust category size effects found in studies such as Landauer and Freedman (1968), Collins and Quillian (1970) and Meyer and Ellis (1970), which were overviewed in Chapter 3. Although the present study did replicate a significant difference in response times between small and medium sized concepts (with sentences with small concepts producing faster reaction times) it was also expected that sentences with large concepts would produce a significantly slower reaction time than sentences with small concepts.

It is seen that the category size effects were weak throughout the studies and may thus highlight the problematic and perhaps idiosyncratic nature of concept size. Not all concepts are going to have the same effect on peoples' memories. It is very dependent on a person's experience and knowledge structure. For instance, a geologist may encounter a rock and this could activate a large semantic area in memory, whereas in the case of a lay person to geology, the rock might only represent a very small and generalised concept. Interestingly though, the s-l sized concept pairings (m-l/s-l condition) in Experiment 4 produced the slowest sentence verification response times and also created the only condition in which there was no significant amount of matching bias in the equivalent conditional inference task. It would therefore seem that when the materials did affect the sentence verification task in line with the predicted category size manipulation, the same materials also eradicated matching bias in the conditional inference task. So, when concept size affects the sentence verification task, it also affects matching bias in the conditional inference task.

Matching Bias on the Sentence Verification Task

It was noted in Chapter 3 that the sentence verification task may be eliciting matching bias. The results from Experiment 4 confirmed this in that there was a main effect of *reference*. True affirmative (TA) and false negative (FN) sentences elicited faster response times than true negative (TN) and false affirmative (FA) sentences. The former two sentence types (TA & FN) contain two concepts that match whereas the latter two contain items that mismatch. Bearing this in mind together with the category size effects previously outlined, it can be seen that there may be real similarities between the sentence verification and the conditional inference task. If there are real similarities between the two tasks then it would show that there may not be a great deal of System2 processing taking place in the conditional inference task. It could be argued that there is very little requirement for the formal and explicit thinking processes of System2 in the sentence verification task. The sentence verification task is one of perceptual processing and fairly instant recognition of the truth or falsity of the sentences. The only role of formal thought processes in this task would be in deciding what is meant by the presence of a negation, which could explain why negated sentences are slower to respond to than non-negated sentences. The conditional inference task by comparison may therefore be relying solely or at least heavily, on unconscious System1 processes (as in the verification task) in order for similar effects of the materials (matching bias and category size effects) to be taking place.

Possible Alternative Factors to Concept Size

Semantic Distance

Collins and Quillian (1969) proposed that memory was organised in a semantic hierarchy (see Chapter 2) such that concepts were stored under superordinate category levels, which were in turn stored under larger categories, thus creating an interconnected framework of concepts and attributes. They have explained the category size effects outlined in Chapter 3 in terms of semantic distance i.e. it takes longer to recognise that a 'collie' is an animal (large category) than it does to ascertain if a 'collie' is a dog (small category). They believe that by using the larger category, an extra step is involved in which it has to be realised that a collie is a dog and that a dog is an animal. This extra step therefore creates longer processing times.

In Experiments 4 and 5, there was found to be mild support for this notion. It will be remembered that in both experiments three sets of word pairs were used; large/medium, large/small and medium/small. It was hypothesised that the 'large/medium' condition would evoke the slowest response times in the verification task and the least amount of endorsements (less matching bias) in the conditional inference task. The hierarchical model of Collins and Quillian suggests that the 'large/small' pairings instead would create the slowest response times in the verification task because of the extra interim step needed to perceive the connection between the two concepts.

It was seen in Experiment 4 that the large/small conditions produced the slowest overall response times on the sentence verification task. Also, in the same experiment the same materials used in the conditional inference task created the only set of inferences in which matching bias was not present. This would offer some support to the semantic distance theory proposed by Collins and Quillian but such a finding though was not subsequently replicated in any of the conditional inference experiments (Exps.5 & 6). The replicated findings suggest stronger support for the concept size manipulation rather than semantic distance approach.

Word Frequency/Familiarity

It was possible that results of the studies may have been caused by participants' familiarity or general everyday exposure to the stimuli rather than the category size manipulation. For instance, it may have been that small concept pairings such as 'oak/tree' were more familiar and easily comprehensible than 'oak/plant' and that this level of comprehension could have affected response times and actual responses. If this were the case, then the fundamental assumptions in the mental activation model would have been incorrect. Experiment 7 set out to investigate this notion by manipulating word frequency/familiarity without manipulating concept size. It was found that the pattern of results obtained previously by manipulating concept size was not replicated. The results of Experiment 7 supported the contention that concept size and not familiarity of the materials were responsible for the reduction in matching bias and therefore offered further support for the mental activation model.

Conclusion

The present study has conclusively shown that the matching bias phenomenon occurs in conditional reasoning problems without the need for lexically matching concepts. This finding has added greater clarity to a previously uncertain area. A new model of reasoning which offers a clearly defined account of unconscious System1 perceptual processes has been put forward. Based almost entirely on proposed underlying mechanisms that may account for memory priming phenomenon, the mental activation model predicts the general findings of the selection task and conditional inference task with hardly any recourse to System2 formal reasoning processes. The mental activation account almost entirely predicts general findings and biases by solely focussing on the materials used in the problems. That is not to say that formal, conscious System2 thought processes are not involved in reasoning, but that the materials used in a problem (especially an abstract problem) have a huge impact on the patterns of responses found from the tasks at a System1 level. This impact is probably greatest when there is uncertainty as to the correct response or confusion concerning the demands of the task.

The mental activation model predicts 'partial matching' on all conditional reasoning tasks (similar to Oaksford's, 2002, account of *Matching*$_2$). This crucial aspect to the present model accounts for differences in response proportions between explicit/implicit affirmation/denial type inferences which are not accounted for by the present matching/mismatching distinction (other mechanisms have to be specified in addition). The mental activation model is thus a parsimonious description of all possible biasing effects of the materials.

In support of the mental activation model, surprising predictions concerning the similarity of materials in the selection task and the size of concepts in the conditional inference task were generated. It has been seen that the results from the present line of studies offer support for these predictions – and which can not readily be addressed by current theories. There has also been some questions raised concerning the notion of relevance. Do card choices in the selection task represent relevance perceptions or are they confounded with extraneous task demands such as the pragmatic inference that the cards selected should verify the premise? Also, there was the finding using indeterminate inference task problems that endorsement responses may not indicate relevance perceptions. It was argued that the term *relevance* may actually be describing *facilitation* in solving the problem and that matching bias may be elicited through refutations of the inference as well as endorsements.

Although the effects of category size may be weak, they are still in the predicted direction. It could be that the abstract nature of the tasks and the focus of the instructions on deductive reasoning might cause participants to consciously attempt to solve the problems using System2 processing. In which case, System1 representational biases may not have as strong an effect were the problems more true to life as in thematic tasks. Thematic tasks are prone to pragmatic influences as were shown in the first chapter. Participants' responses can be systematically altered depending on simple changes of emphasis of the problems. If the concept size manipulation of memory can affect responses at these current basic levels involving abstract tasks, then it is only reasonable that changes or spreads of activation will also play a part in thematic tasks.

It was pointed out in Chapter 1 that a possible biasing/debiasing effect of thematic materials may be due to 'enrichment'. Enrichment in this case points to more mental associations. In a distributed semantic network, applying the principles of a finite amount of activation proposed by Anderson (1989) would in the case of enrichment, involve a spread of activation to a wider semantic area. It was put forward that such a spread of activation could reduce the perceived relevance of some problem concepts thus creating more uniform and equivalent intuitions of relevance. Using large concepts could be seen as a form of enrichment in that there are more semantic associations connected to larger concepts. Thus, there would be a reduced biasing effect of the materials (none in particular would stand out in the initial representation of the problem) which would enable System2 processes to work on more elaborate forms of the problem and reduce representational System1 biases such as matching. Only by understanding biases at these basic, perceptual and abstract levels will any clarity be added to the findings concerning the pragmatic influences of thematic, real world reasoning tasks.

References

Anderson, J.R. (1976). Language, memory and thought. Hillsdale: Erlbaum.

Anderson, J.R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Anderson, J.R. (1991). The adaptive nature of human categorization. Psychological Review, 98(3), 409-429.

Anderson, J.R. & Milson, R. (1989). Human memory: An adaptive perspective. Psychological Review, 96(4), 703-719.

Becker, S., Moscovitch, M., Behrmann, M. & Joordens, S. (1997). Long-term semantic priming: A computational account and empirical evidence. Journal of Experimental Psychology: Learning, Memory and Cognition, 23(5), 1059-1082.

Bentin, S. & Feldman, L. (1990). The contribution of morphologic and semantic relatedness to repetition priming at short and long lags: Evidence from Hebrew. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 42(A), 693-711.

Bowers, J.S. (2000). In defence of abstractionist theories of word identification and repetition priming. Psychonomic Bulletin and Review, 7, 83-99.

Brown, W.P. (1972). Studies in word listing: Some norms and their reliability. Irish Journal of Psychology, 1(3), 117-159.

Bruce, V. & Valentine, T. (1985). Identity priming in the recognition of familiar faces. British Journal of Psychology, 76, 373-383.

Chang, T.M. (1986). Semantic memory: Facts and models. Psychological Bulletin, 99(2), 199-220.

Chater, N. (1995). Neural networks: the new statistical models of the mind. In J.P.Levy, D.Bairaktaris, J.A.Bullinaria, P.Cairns (Eds.), Connectionist models of memory and language. London: UCL Press.

Chater, N. (1999). The search for simplicity: A fundamental cognitive principle? The Quarterly Journal of Experimental Psychology, 52A(2), 273-302.

Cheng, P.W., & Holyoak, K.J. (1985). Pragmatic reasoning schemas. Cognitive Psychology, 17, 391-416.

Collins, A.M. & Loftus, E.F. (1975). A spreading activation theory of semantic processing. Psychological Review, 82, 407-428.

Collins, A.M. & Quillian, M.R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behaviour, 8, 240-247.

Collins, A.M. & Quillian, M.R. (1970). Does category size affect categorization time? . Journal of Verbal Learning and Verbal Behaviour, 9, 432-438.

Dennis, I. & Schmidt, K. (In press). Associative processes in repetition priming. Journal of Experimental Psychology: Learning, Memory and Cognition.

Evans, J.St.B.T. (1972). Interpretation and matching bias in a reasoning task. Quarterly Journal of Experimental Psychology, 24, 193-199.

Evans, J.St.B.T. (1975). On interpreting reasoning data. A reply to Van Duyne. Cognition, 3, 387-390.

Evans, J.St.B.T. (1982). The psychology of deductive reasoning. London: Routledge & Kegan Paul.

Evans, J.St.B.T. (1984). Heuristic and analytic processes in reasoning. British Journal of Psychology, 75,451-468.

Evans, J.St.B.T. (1989). Bias in human reasoning: Causes and consequences. Hove: Erlbaum.

Evans, J.St.B.T. (1991). Theories of human reasoning: The fragmented state of the art. Theory & Psychology, 1, 83-105.

Evans, J.St.B.T. (1993). The mental model theory of conditional reasoning: Critical appraisal and revision. Cognition, 48, 1-20.

Evans, J.St.B.T. (1995). Relevance and reasoning. In S.E.Newstead & J.St.B.T.Evans (Eds.) Perspectives on thinking and reasoning. Hove: Erlbaum.

Evans, J.St.B.T. (1998) Matching bias in conditional reasoning: Do we understand it after 25 years. Thinking and Reasoning, 4(1), 45-82.

Evans, J.St.B.T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. Psychological Bulletin, 128(6), 978-996.

Evans, J.St.B.T., & Clibbens, J. (1995).. Perspective shifts in the selection task: Reasoning or relevance? Thinking and Reasoning, 1, 315-323.

Evans, J.St.B.T., Clibbens, J. & Rood, B. (1995). Bias in conditional inference: Implications for mental models and mental logic. The Quarterly Journal of Experimental; Psychology, 48A(3), 644-670.

Evans, J.St.B.T., Clibbens, J. & Rood, B. (1996). The role of implicit and explicit negation in conditional reasoning bias. Journal of Memory and Language, 35, 392-409.

Evans, J.St.B.T. & Handley, S.J. (1999). The role of negation in conditional inference. The Quarterly Journal of Experimental Psychology, 52A(3), 739-769.

Evans, J.St.B.T., Legrenzi, P., & Girroto, V. (1997). The influence of linguistic context on propositional reasoning: The case of matching bias. Unpublished manuscript, University of Plymouth, UK.

Evans, J.St.B.T. & Lynch, J.S. (1972). Matching bias in the selection task. British Journal of Psychology, 64(3), 391-397.

Evans, J.St.B.T., & Newstead, S.E. (1980). A study of disjunctive reasoning. Psychological Research, 41, 373-388.

Evans, J.St.B.T., & Newstead, S.E. & Byrne, R.M.J. (1993). Human reasoning: The psychology of deduction. Hove: Erlbaum.

Evans, J.St.B.T., & Over, D. E. (1996). Rationality and reasoning. Hove: Psychology Press.

Evans, J.St.B.T. & Over, D.E. (1999). Explicit representations in hypothetical thinking. Behavioural and Brain Sciences, 22, 763-764.

Feustel, T.C., Schiffrin, R.M. & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. Journal of Experimental Psychology: General, 112, 309-346.

Forbach, G.B., Stanners, R.F. & Hochhaus, L. (1974). Repetition and practice effects in a lexical decision task. Memory and Cognition, 2, 337-339.

Graf, P., Mandler, G. & Haden, P. (1982). Simulating amnesic symptoms in normal subjects. Science, 218, 1243-1244.

Graf, P. & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. Journal of Verbal Learning and Verbal Behaviour, 23, 553-568.

Graf, P. and Masson, M.E.J. (1993). Implicit memory: New directions in cognition, development and neuropsychology.

Griggs, R.A & Cox, J.R. (1982). The elusive thematic materials effect in Wason's selection task. British Journal of Psychology, 73, 407-420.

Groeger, J.A. (1988). Qualitatively different effects of undetected and unidentified auditory primes. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 40A(2), 323-339.

Hinton, G.E., McClelland, J.L. & Rumelhart, D.E. (1986). Distributed representations. In D.E. Rumelhart & J.L. McClelland (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1, (77-109). Cambridge, MA: MIT Press

Holyoak, K.J., & Cheng, P.W. (1995). Pragmatic reasoning with a point of view. Thinking and Reasoning, 1, 289-313.

Jacoby, L.L. (1983). Perceptual enhancement: Persistent effects of an experience. Journal of Experimental Psychology: Learning, Memory and Cognition, 9, 21-38.

Jacoby, L.L. & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. Journal of Experimental Psychology: General, 110, 306-340.

Johnson-Laird, P.N. (1983). Mental models. Cambridge: Cambridge University Press.

Johnson-Laird, P.N. & Byrne, R.M.J. (1991). Deduction. Hove: Erlbaum

Johnson-Laird, P.N. & Byrne, R.M.J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. Psychological Review, 109(4), 646-678.

Johnson-Laird, P.N., Legrenzi, P., & Legrenzi, M.S. (1972). Reasoning and a sense of reality. British Journal of Psychology, 63, 395-400.

Johnson-Laird, P.N. & Tagart, J. (1969). How implication is understood. American Journal of Psychology, 2, 367-373.

Joordens, S. & Becker, S. (1997). The long and short of semantic priming effects in lexical decision. Journal of Experimental Psychology: Learning, Memory and Cognition, 23(5), 1083-1105.

Kirby, K.N. (1994). Probabilities and utilities of fictional outcomes in Wason's four-card selection task. Cognition, 51, 1-28.

Kolers, P.A. (1975). Memorial consequences of automatized encoding. Journal of Experimental Psychology: Human Learning and Memory, 1, 689-701.

Kolers, P.A. (1976). Reading a year later. Journal of Experimental Psychology: Human Learning and Memory, 2, 554-565.

Landauer, T.K. & Freedman, J.L. (1968). Information retrieval from long-term memory: Category size and recognition time. Journal of Verbal Learning and Verbal Behaviour, 7, 291-295.

Landauer, T.K. & Meyer, D.E. (1972). Category size and semantic-memory retrieval. Journal of Verbal Learning and Verbal Behaviour, 11, 539-549.

Lea, R.B. & Mulligan, E.J. (2002). The effect of negation on deductive inferences. Journal of Experimental Psychology: Learning, Memory and Cognition, 28(2), 303-317.

Legrenzi, P., Girotto, V. & Johnson-Laird, P.N. (1993). Focussing in reasoning and decision making. Cognition, 49, 37-66.

Loftus, E.F. & Scheff, R.W. (1971). Categorization norms for 50 representative instances. Journal of Experimental Psychology, 91, 355-364.

MacDonald, M.C. & Just, M.A. (1989). Changes in activation levels with negation. Journal of Experimental Psychology: Learning, Memory and Cognition, 15, 633-642.

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

Markovits, H., Doyon, C. & Simoneau, M. (2002). Individual differences in working memory and conditional reasoning with concrete and abstract content. Thinking and Reasoning, 8(2), 97-107.

Masson, M.E.J. (1984). Memory for the surface structure of sentences: Remembering with and without awareness. Journal of Verbal Learning and Verbal Behaviour, 23, 579-592.

Meyer, D.E. & Ellis, G.B. (1970). Parallel processes in word recognition. Paper presented at the meeting of the Psychonomic Society, San Antonio, November.

McClelland, J.L. & Rumelhart, D.E. (1985). Distributed memory and the representation of general and specific information. Journal of Experimental Psychology: General, 114, 159-188.

McClelland, J.L. & Rumelhart, D.E. (1986). A distributed model of human learning and memory. In J.L. McClelland & D.E.Rumelhart (Eds.), Parallel distributed processing Vol. 2 (170-215). Cambridge, MA: MIT Press.

McCloskey, M. (1980). The stimulus familiarity problem in semantic research. Journal of Verbal Learning and Verbal Behaviour, 19, 485-502.

McNamara, T.P. (1992). Theories of priming: I. Associative distance and lag. Journal of Experimental Psychology: Learning, Memory and Cognition, 18(6), 1173-1190.

McNamara, T.P. (1994). Theories of priming: II. Types of primes. Journal of Experimental Psychology: Learning, Memory and Cognition, 20(3), 507-520.

Morris, C.D., Bransford, J.D. & Franks, J.J. (1977). Levels of processing versus transfer appropriate processing. Journal of Verbal Learning and Verbal Behaviour, 16(5), 519-533.

Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In P.A. Kolers, M.E. Wrolstad & H. Bouma (Eds.), Processing models of visible language. New York: Plenum.

Oaksford, M. (2002). Contrast classes and matching bias as explanations of the effects of negation on conditional reasoning. Thinking and Reasoning, 8(2), 135-151.

Oaksford, M. & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. Psychological Review, 101, 608-631.

Oaksford, M., Chater, N., & Grainger, B. (1999). Probabilistic effects in data selection. Thinking and Reasoning, 5, 193-243.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. Journal of Experimental Psychology: Learning, Memory & Cognition, 26, 883-899.

Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. Journal of Experimental Psychology: Learning, Memory and Cognition, 18, 835-854.

Pollard, P. & Evans, J.St.B.T. (1987). Content and context effects in reasoning. American Journal of Psychology, 100(1), 41-60.

Reber, A.S. (1993). Implicit Learning and Tacit Knowledge. Oxford: Oxford University Press.

Richardson-Klavehn, A., & Bjork, R.A. (1988). Measures of memory. Annual Review of Psychology, 39, 475-543.

Roediger, H.L. (1990). Implicit memory: Retention without remembering. American Psychologist, 45, 1043-1056.

Roediger, H.L.III & Srivinas, K. (1993). Specificity of operations in perceptual priming. In P. Graf & M.E.J. Masson (Eds.) Implicit memory: New directions in cognition, development and neuropsychology (17-48). Hillsdale, NJ: Erlbaum.

Rumelhart, D.E., Hinton, G.E. & McClelland, J.L. (1986). A general framework for parallel distributed processing. In D.E. Rumelhart & J.L. McClelland (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1, (45-76). Cambridge, MA: MIT Press.

Rumelhart, D.E., Smolensky, P., McClelland, J.L. & Hinton, G.E. (1986). Schemata and sequential thought processes in PDP models. In J.L. McClelland & D.E. Rumelhart (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 2, (7-57). Cambridge, MA: MIT Press.

Scarborough, D.L., Gerard, L. & Cortese, C. (1979) Accessing lexical memory: The transfer of word repetition effects across task and modality. Memory and Cognition, 7, 3-12.

Schacter, D.L. (1987). Implicit memory: History and current status. Journal of Experimental Psychology: Learning, Memory and Cognition, 13(3), 501-518.

Schacter, D.L & Graf, P. (1986). Effects of elaborative processing on implicit and explicit memory for new associations. Journal of Experimental Psychology: Learning, Memory and Cognition, 12, 432-444.

Schroyens, W.J., Schaeken, W. & d'Ydewalle, G. (2001). The processing of negations in conditional reasoning: A meta-analytic case study in mental model and/or mental logic theory. Thinking & Reasoning, 7(2), 121-172.

Schroyens, W., Verschueren, N., Schaeken, W. & d'Ydewalle, G. (2000). Conditional reasoning with negations: Implicit and explicit affirmation or denial and the role of contrast classes. Thinking and Reasoning, 6(3), 221-251.

Sloman, S.A. (1996). The empirical case for two systems of reasoning. Psychological Bulletin, 119, 3-22.

Smith, E.S., Shoben, E.J. & Rips, L.J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. Psychological Review, 81(3), 214-241.

Sperber, D., Cara, F. & Girotto, V. (1995). Relevance theory explains the selection task. Cognition, 57, 31-95.

Squire, L.R. (1986) Mechanisms of memory. Science, 232, 1612-1619.

Stanovich, K.E. (1999). Who is rational? Studies of individual differences in reasoning. Mahway: Erlbaum.

Stanovich, K.E. & West, R.F. (2000). Individual differences in reasoning: Implications for the rationality debate. Behavioural and Brain Sciences, 23(5), 645-726.

Stark, C.E.L. & McClelland, J.L. (2000). Repetition priming of words, pseudowords and non-words. Journal of Experimental Psychology: Learning, Memory and Cognition, 26 (4), 945-972.

Tenpenny, P.L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. Psychonomic Bulletin and Review, 2, 339-363.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press.

Tulving, E. (1983). Elements of episodic memory. Oxford: Oxford University Press.

Tulving, E., Schacter, D.L. & Stark, H.A. (1982). Priming effects in word fragment completion are independent of recognition memory. Journal of Experimental Psychology: Learning, Memory and Cognition, 8, 336-342.

Vriezen, E.R., Moscovitch, M. and Bellos, S.A. (1995). Priming effects in semantic classification tasks. Journal of Experimental Psychology: Learning Memory and Cognition, 21, 933-946.

Wason, P.C. (1959). The processing of positive and negative information. Quarterly Journal of Experimental Psychology, 11, 92-107.

Wason, P.C. (1961). Response to affirmative and negative binary statements. British Journal of Psychology, 52(2), 133-142.

Wason, P.C. (1965). The context of plausible denial. Journal of Verbal Learning and Verbal Behaviour, 4, 7-11.

Wason, P.C. (1966). Reasoning. In B.M.Foss (Ed.), New horizons in psychology. Harmondsworth: Penguin.

Wason, P.C. (1968). On the failure to eliminate hypotheses – A second look. In P.C. Wason & P.N.Johnson-Laird (Eds), Thinking and Reasoning. Harmondsworth: Penguin.

Williamsen, J.A., Johnson, H.J. & Eriksen, C.W. (1965). Some characteristics of posthypnotic amnesia. Journal of Abnormal Psychology, 70, 123-131.

Yama, H. (2001). Matching versus optimal data selection in the Wason selection task. Thinking and Reasoning, 7, 295-311.

Young, A.W., McWeeny, K.H., Hay, D.C. & Ellis, A.W. (1986). Access to identity-specific semantic codes from familiar faces. Quarterly Journal of Experimental Psychology, 38A, 271-295.

Appendices

Appendix A

All 16 negations-paradigm inferences (implicit negation)

If there is an A on the left side then there is a 2 on the right side. **(MP1)**
There is an A on the left side.
Therefore there is a 2 on the right side.

If there is an A on the left side then there is a 2 on the right side. **(AC1)**
There is a 2 on the right side.
Therefore there is an A on the left side.

If there is an A on the left side then there is a 2 on the right side. **(MT1)**
There is a 3 on the right side.
Therefore there is NOT an A on the left side.

If there is an A on the left side then there is a 2 on the right side. **(DA1)**
There is a B on the left side.
Therefore there is NOT a 2 on the right side.

If there is an A on the left side then there is NOT a 2 on the right side. **(MP2)**
There is an A on the left side.
Therefore there is NOT a 2 on the right side.

If there is an A on the left side then there is NOT a 2 on the right side. **(AC2)**
There is a 3 on the right side.
Therefore there is an A on the left side.

If there is an A on the left side then there is NOT a 2 on the right side. **(MT2)**
There is a 2 on the right side.
Therefore there is NOT an A on the left side.

If there is an A on the left side then there is NOT a 2 on the right side. **(DA2)**
There is a B on the left side.
Therefore there is a 2 on the right side.

Continued overleaf

If there is NOT an A on the left side then there is a 2 on the right side. **(MP3)**
There is a B on the left side.
Therefore there is a 2 on the right side.

If there is NOT an A on the left side then there is a 2 on the right side. **(AC3)**
There is a 2 on the right side.
Therefore there is NOT an A on the left side.

If there is NOT an A on the left side then there is a 2 on the right side. **(MT3)**
There is a 3 on the right side.
Therefore there is an A on the left side.

If there is NOT an A on the left side then there is a 2 on the right side. **(DA3)**
There is an A on the left side.
Therefore there is NOT a 2 on the right side.

If there is NOT an A on the left side then there is NOT a 2 on the right side. **(MP4)**
There is a B on the left side.
Therefore there is NOT a 2 on the right side.

If there is NOT an A on the left side then there is NOT a 2 on the right side. **(AC4)**
There is a 3 on the right side.
Therefore there is NOT an A on the left side.

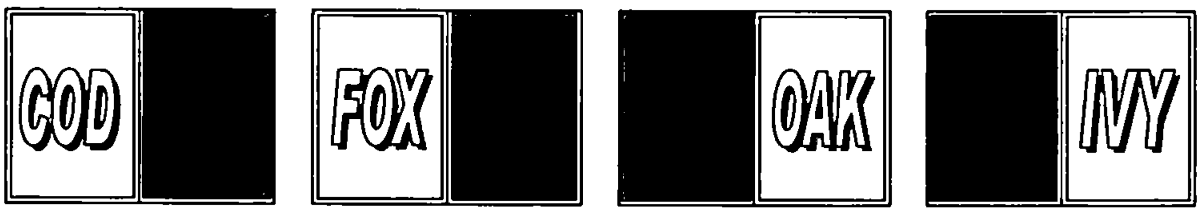
If there is NOT an A on the left side then there is NOT a 2 on the right side. **(MT4)**
There is a 2 on the right side.
Therefore there is an A on the left side.

If there is NOT an A on the left side then there is NOT a 2 on the right side. **(DA4)**
There is an A on the left side.
Therefore there is a 2 on the right side.

<u>Appendix B</u>

Materials Used in Experiment 1

Example of 'Semantic' Problem Layout



Each of the four cards below has an animal on the left side and a plant on the right side.

The following rule applies to these four cards and may be true or false:

"If a card has a fish on the left side, then it has a tree on the right side."

Which card or cards would you need to unmask in order to discover whether or not the above rule is true?

Materials

Lexical Condition Materials

Scenario Category	Scenario Category	Premise Antecedent	Premise Consequent	Cards Displayed			
animal	plant	**cod**	**oak**	cod	fox	oak	ivy
city	building	**London**	**Tate**	London	Exeter	Tate	Big Ben
food	beverage	**eclair**	**beer**	eclair	pasta	beer	coke
transport	entertainment	**Escort**	**CD**	Escort	747	CD	film

Semantic Condition Materials

Scenario Category	Scenario Category	Premise Antecedent	Premise Consequent	Cards Displayed			
animal	plant	**fish**	**tree**	cod	fox	oak	ivy
city	building	**capital**	**gallery**	London	Exeter	Tate	Big Ben
food	beverage	**cake**	**alcohol**	eclair	pasta	beer	coke
transport	entertainment	**car**	**music**	Escort	747	CD	film

Appendix C

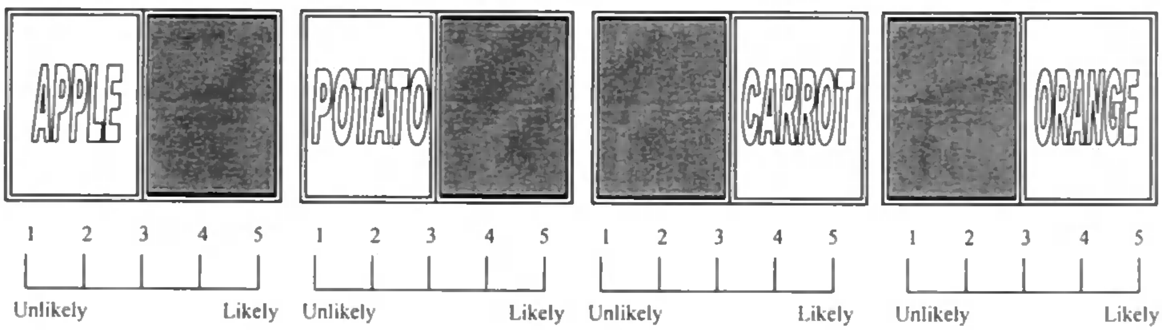
Materials Used in Experiment 3

Example of '*Similar*' Problem Layout:

Each of the four cards below has one item on the left side and one item on the right side.

The following rule applies to these four cards and may be true or false:

"If a card has a fruit on the left side, then it has a vegetable on the right side."



- Mark each scale as to how likely you think that card could falsify the rule if it were unmasked.
- Circle the card or cards you would need to unmask in order to discover whether or not the above rule is true or false?

Materials

Dissimilar Condition Materials

Premise Antecedent	Premise Consequent	Cards Displayed			
food	**drink**	cake	ivy	beer	shoe
fruit	**vegetable**	apple	glass	carrot	fox
city	**country**	London	pot	France	knee
tree	**flower**	oak	brick	pansy	train

Similar Condition Materials

Premise Antecedent	Premise Consequent	Cards Displayed			
food	**drink**	cake	tea	beer	bread
fruit	**vegetable**	apple	potato	carrot	orange
city	**country**	London	England	France	Paris
tree	**flower**	oak	rose	pansy	willow

NB Differences between conditions underlined

Appendix D

Materials Used in Experiment 4

Six 'nested triple' categories were used for both the sentence verification task and the conditional inference task: *animal/reptile/adder, plant/tree/oak, food/cake/eclair, mineral/gem/diamond, activity/sport/tennis, building/gallery/the Tate.*

Only two instances from each 'triple were used in any one task. These pairings formed the M-l, S-l, and S-m concept-size conditions:

M-l	reptile	tree	cake	gem	sport	gallery
	animal	plant	food	mineral	activity	building
S-l	adder	oak	eclair	diamond	tennis	the Tate
	animal	plant	food	mineral	activity	building
S-m	adder	oak	eclair	diamond	tennis	the Tate
	reptile	tree	cake	gem	sport	gallery

- For the sentence verification task, the first instance in the pair would be the first instance in the sentence e.g. (M-l condition) 'a *reptile* is an *animal'*.
- For the conditional inference task, the first instance in the pair would be the categorical premise e.g. (M-l condition) 'If there is an *animal* on the left there is a cake on the right. There is a *reptile* on the left'.

Appendix E

Materials Used in Experiment 5

The 24 nested triples.

M-l Concepts Pairs:

Large Concept	Medium Concept	Small Concept
Mineral	Gem	*Diamond*
Food	Cake	*Eclair*
Plant	Tree	*Oak*
Building	Gallery	*the Tate*
Animal	Reptile	*Adder*
Living thing	Mammal	*monkey*
Activity	Sport	*Tennis*
Clothing	Footwear	*Boot*

S-l Concepts Pairs:

Large Concept	Medium Concept	Small Concept
Beverage	*Wine*	Chardonnay
Transport	*Aeroplane*	747
Occupation	*Doctor*	Paediatrician
Manmade object	*Furniture*	Chair
Place	*Country*	France
Landscape feature	*Mountain*	Everest
Degree course	*Science*	Chemistry
Vehicle	*Car*	Vauxhall

S-m Concept Pairs:

Large Concept	Medium Concept	Small Concept
Structure	Dwelling	House
Electrical conductor	Metal	Copper
Implement	Tool	Hammer
Artform	Picture	Photograph
Colour	Red	Crimson
Person	Relative	Uncle
Creature	Insect	Ant
Liquid	Drink	Water

The items in italics were not used in each condition as only *pairs* of items were used. They are included to highlight how the terms formed nested triple categories.

Appendix F

Materials Used in Experiment 6

The 8 nested triples.

Large Concept	Medium Concept	Small Concept
Mineral	Gem	Diamond
Food	Fruit	Apple
Plant	Flower	Rose
Creature	Insect	Ant
Animal	Reptile	Adder
Liquid	Drink	Water
Activity	Sport	Tennis
Place	Country	France

Two pairs of concepts from each 'triple' were used for each of the three *concept pairings*:

M-l (medium concept * large concept)

S-l (small concept * large concept)

S-m (small concept * medium concept)

For each concept paring, the first named (largest) instance formed the *categorical premise* e.g (M-l condition):

'If there is NOT a <u>gem</u> on the left side there is a plant on the right side. There is NOT a <u>mineral</u> on the left side.'

NB. Only 'implicit inferences' were used in this experiment and therefore the concept sizes were reversed for the categorical premise (largest) /referred clause pairing (smallest). This is the opposite way round to the concept size placement in Experiments 4 and 5.

Appendix G

Materials Used in Experiment 7

Categories with infrequent and frequent instances taken from Brown's (1972) word norms.

Category	Infrequent Concept	Frequent Concept
Animal	Reindeer	Dog
Bird	Turkey	Sparrow
Item of clothing	Corset	Jacket
Drink	Ale	Whisky
Musical instrument	Bugle	Piano
Precious stone	Garnet	Diamond
Sport	Archery	Football
Female name	Maria	Anne

Inferences were either formed with an infrequent or frequent instance forming the categorical premise e.g.

'If there is an <u>animal</u> on the left then there is a drink on the right. There is a <u>reindeer</u> <u>(dog)</u> on the left.'

NB The parenthesis indicates the concept shown in the *frequent* group.