Australian Government
Department of Defence
Defence Science and
Technology Organisation

Linear Separability in Categorisation and Inference: A Test of the Johnson-Laird Falsity Model

Susannah J. Whitney, George Galanis and Armando Vozzo

Land Operations Division
Defence Science and Technology Organisation

DSTO-TR-2935

ABSTRACT

Johnson-Laird suggests that difficulties in problem solving can be explained by the mental models theory. This study tests linear seperability effects in categorisation and inference as an alternate explanation, hypothesising that categorisation and inference would be easier for linearly separable (LS) functions than nonlinearly separable (NLS). Thirty two participants were tested on one LS and one NLS function over repeated trials. Results indicated that categorisation and inference were significantly more difficult for NLS functions, but only for the highest performing participants on some trials. Among poorer performing participants there were no significant differences between response rates and response times. The most likely explanations for these findings are the complexity and duration of the experiment, rather than lack of support for the linear separability hypothesis. Implications for the military and research communities and suggestions for future research are discussed.

RELEASE LIMITATION

Approved for public release

APPROVED FOR PUBLIC RELEASE

Linear Separability in Categorisation and Inference: A Test of the Johnson-Laird Falsity Model

Executive Summary

In cognitive science, categorisation refers to the ability to use a set of characteristics to determine which category an object belongs to. Inference is the ability, given category membership and some defining characteristics, to deduce the values of other characteristics. These processes can be difficult when categorisation rules are complex, or when multiple characteristics need to be considered. In addition, the difficulty of categorisation can be affected by linear separability; that is, the extent to which category membership is tightly clustered, or more loosely bound.

DSTO researchers have suggested that linear separability may explain a common effect in cognitive psychology; the tendency for people to incorrectly answer problems such as:

> Only one statement about a hand of cards is true:
> (1). There is a King or Ace or both
> (2). There is a Queen or Ace or both
> Which is more likely, King or Ace?

While it is intuitive to answer 'Ace', as it occurs in both statements, the correct answer is 'King'. As only one statement can be true, the Ace can logically never occur, since its presence makes both statements true.

The prominent theory for the difficulties people encounter in solving problems like the example above is Johnson-Laird's mental model theory, which suggests that people construct incomplete models of all possible answers. However, in this paper, we test an alternative explanation, that of linear separability. This explanation predicts that problems will be easier to solve when they are linearly separable (LS); that is, when it is straightforward to separate correct from incorrect answers. In contrast, nonlinearly separable (NLS) problems, where it is more complex to separate correct from incorrect answers, will be more difficult to solve.

To test this hypothesis, 32 military and civilian participants completed an experiment. Participants were informed that there was a hypothetical light switch, which was controlled by three switches that could be on or off. The purpose of the experiment was

to learn and understand the rule that determined whether or not the light was on. In the Categorisation Phase, participants were presented with the eight possible combinations of the three switches, and asked to judge if the light was on and off. Light switch combinations were displayed one at a time, and participants were given immediate feedback on their decision. After eight presentations of the eight combinations, participants were given an inference test, comprising between seven and nine questions. They were shown one or two of the switches, and the light state (on or off), and asked what could be deduced about another switch. This categorisation and inference sequence was repeated five times. Categorisation tests were repeated, but each inference test was unique. Each participant was tested on one LS and one NLS function, randomly selected from a pool of five LS and five NLS functions.

Overall, results showed no significant differences between response rates and response times for categorisation and inference of LS and NLS functions. However, when analyses were confined to the highest performing participants, some significant differences were found between LS and NLS functions during categorisation and inference phases. This suggests that the experiment in its current form may have been too difficult and too long for participants to remain engaged and to understand the experimental requirements. However, the linear separability explanation is still plausible, and should be further investigated.

This work was conducted under the Enabling Research Program (ERP) of the Land Operations Division (LOD) Land Human Sciences Major Science and Technology Capability (LHS MSTC). (Land Operations Division was subsequently renamed Land Division in the 2013 DSTO restructure). By definition, the ERP includes work that:

- has the potential for a high payoff in the medium to long term that addresses a need important to the Australian Defence land environment, and
- is not part of the LOD client program, and will probably not be supported by the client in the near term.

The potential for high payoff in this work is through investigating fundamental issues surrounding decision making. Military decision-makers have to make decisions under pressure with information constraints. Hence, while this work was not part of the LOD client program, it was the potential to, in the long term, help facilitate the development and improvement of military decision making, including support systems.

Authors

Susannah J. Whitney
Land Division

Susannah Whitney holds a PhD in Psychology from the University of Queensland, and a First Class Honours degree in Psychology from the University of Newcastle. She has been employed as a Cognitive Scientist at DSTO since February 2007. Her research interests include the psychological principles underpinning learning and memory, the training effectiveness of new technology, and skills required for successful training. Susannah is also a member of the DSTO Ethics Review Panel.

_____   _____

George Galanis
Land Division

George Galanis received an A.Mus.A from Melbourne University in 1978, and a B.Sc. (Computer Science) from Monash University in 1979. George joined DSTO in 1991 working in the field of flight simulation, and he had a leading role in the acquisition and development of the Air Operations Simulation Centre at Fisherman's Bend. He was awarded a PhD in 2000 for his thesis on the topic of human perception in the control of aircraft. He then moved to Adelaide to take up the role of Head of Training Technology with the Land Operations Division at DSTO Edinburgh. George is currently the Air Force Scientific Adviser.

_____   _____

Armando Vozzo
Land Division

Armando Vozzo holds a Bachelor of Engineering with Honours and a Graduate Diploma in Adult Education from the University of South Australia, and a Masters degree in Human Factors from the University of Queensland. His research interests relate to the use of simulation for training and the training effectiveness of simulation environments. He has led scientific analysis teams for project evaluations and field studies with Army. Armando also has experience in private industry, working on a wide area surveillance project and lecturing at university. His current role is the Principal Staff Officer (Science) at the Australian Army's Forces Command (FORCOMD).

_____   _____

This page is intentionally blank

Contents

Acronyms

AI	Artificial Intelligence
ERP	Enabling Research Program
LHS	Land Human Sciences
LOD	Land Operations Division
LS	Linearly separable
NLS	Nonlinearly separable
$p\eta^2$	Partial eta-squared (measure of effect size)
RT	Response time
SD	Standard deviation

This page is intentionally blank

1. Introduction

1.1   Overview

Researchers such as Johnson-Laird and his colleagues have consistently demonstrated (Barres and Johnson-Laird, 2003; Goodwin and Johnson-Laird, 2011; Johnson-Laird et al., 2009; Johnson-Laird and Savary, 1996) that people have trouble solving complicated reasoning problems, such as the following:

> Suppose that only one of the following assertions is true:
> (1) You have the mints.
> (2) You have the gum or the lollipops, but not both.
> Also, suppose you have the mints. What, if anything, follows? Is it possible
> that you also have either the gum or the lollipops? Could you have both? [1]
> (Khemlani and Johnson-Laird, 2009)

The difficulty in solving such problems (which we term the "Johnson-Laird effect") has typically been explained by theories of mental models (Johnson-Laird, 2010; Johnson-Laird and Savary, 1996; Johnson-Laird and Savary, 1999). However, we propose an alternate approach which may more accurately explain Johnson-Laird's findings. This approach draws on categorisation and inference research, and suggests that the difficulty in solving these problems is dependent on linear separability. At a simplistic level, linear separability refers to the extent to the degree of similarity between category members, and the extent to which objects can be easily divided into categories. A more complex definition of linear separability is provided in Section 1.3.

This report documents a study testing the linear separability explanation as an alternative explanation for the Johnson-Laird effect It is intended to be read in conjunction with Whitney (2013), which provides additional detail on the previous research in categorisation, inference, linear separability and mental models. This work was sponsored by the Chief, Land Operations Division (LOD), and was conducted under LOD's Land Human Sciences (LHS) Enabling Research Program (ERP)[2].

This study brings together two distinct groups of research in cognitive psychology, firstly, Johnson-Laird's work on mental models, and secondly, the concepts of categorisation, inference, and linear separability. While Johnson-Laird's research paradigm is the focus of the study, categorisation, inference, and linear separability are discussed first. This is because it is important to understand these concepts as they are traditionally applied in cognitive psychology before being able to understand how we apply them to Johnson-Laird's work.

---

[1] An explanation of the correct answer for this problem is given in Section 1.4.
[2] Land Operations Division formally became Land Division under the 2013 DSTO restructure. The divisional names in use at the time the study was conducted are used in this report.

1.2   Categorisation and inference

Categorisation refers to the ability to group objects on the basis of their attributes or characteristics. Inference refers to the ability to use category membership and some attributes to infer the value of other attributes (Yamauchi and Markman, 1998). To illustrate the concepts of categorisation and inference, consider the set of objects in Figure 1. Each object has two characteristics, shape (circle or triangle) and colour (red or blue). They have been divided into two categories, Category A, and Category B. Based on the information in the figure, it appears that category membership is determined by colour. If an object is red, it belongs to Category A, and if it is blue, it belongs to Category B.

Once these category rules are known, the ability to categorise and make inferences can be tested. A categorisation problem might show a novel object, such as a red circle, and ask which category it belonged to. An inference problem might show an object, such as a rectangle, indicate that it belongs to Category B, and ask the likely colour of the object. Using the categorisation rule in Figure 1, solving these problems is straightforward. However, with more complex category membership rules, categorisation and inference become more difficult.
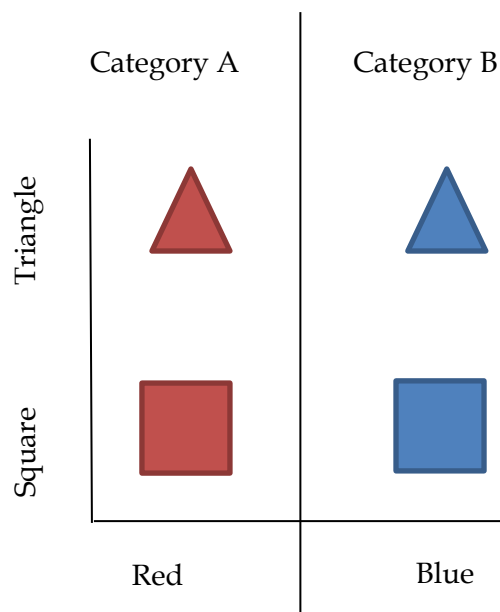
Figure 1: Simple categorisation

Categorisation and inference are important for a number of reasons. They have practical, everyday importance in helping us make decisions (e.g. Is this loaf of bread fresh or stale?) or deductions (e.g. I know my colleague votes for an opposing political party, so I assume our views on a contentious political issue will be different). In addition, understanding the way in which people make categorisation and inference decisions helps contribute to formal theories of the way we acquire, process, and structure information.

1.3   Linear separability

One factor affecting categorisation and inference is linear separability. Categorisation can be either linearly separable (LS) or non-linearly separable (NLS). For categorisation containing two dimensions, as in Figure 1, and the examples given in Figures 2 and 3, categorisation is LS where a single straight line can be drawn in the two dimensional problem space that separates the two categories. It is NLS where the two categories cannot be separated using a single straight line (Blair and Homa, 2001).

To illustrate LS categorisation, consider Figure 2. The objects are the same as in Figure 1, but different rules determine category membership. In Figure 2, Category A comprises objects that are red or a triangle or both, and Category B comprises all other objects[3]. As the figure shows, it is possible to draw a single line separating Category A and Category B, hence this categorisation is LS.
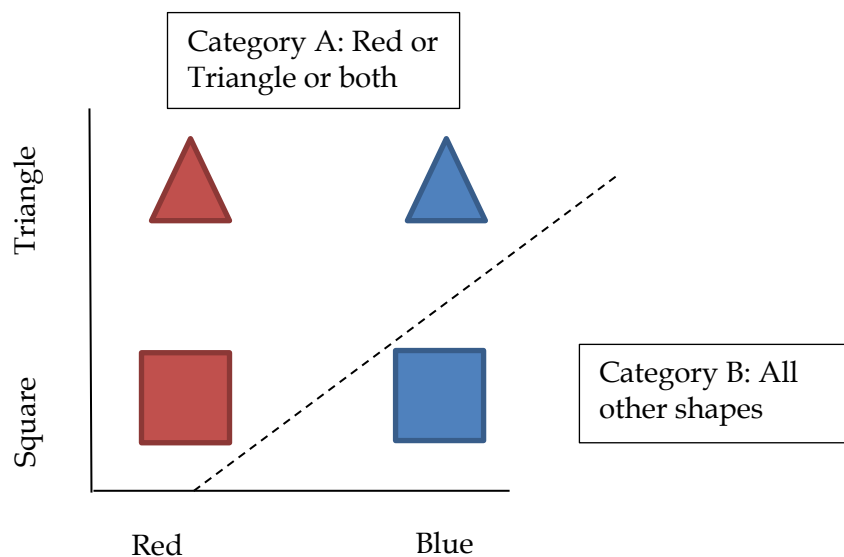

Figure 2: Linearly separable categorisation

Figure 3 shows different categorisation rules for the same objects. Here, an object belongs to Category A if it is blue or a square (but not both blue and a square), and belongs to Category B if it is red or a square (but not both red and a square)[4]. In this example, it is not possible to draw a single line separating the two members of Category A from the two members of Category B. Hence, this is an example of NLS categorisation.

---

[3] For ease of reading, the category memberships in Figure 2 and Figure 3 are explained using plain English. The formal membership rules expressed in Boolean logic are, Category A: (Red OR Triangle), and Category B: NOT (Red OR Triangle).

[4] Formal membership rules are, Category A: (Blue XOR Square), Category B: (Red XOR Square).
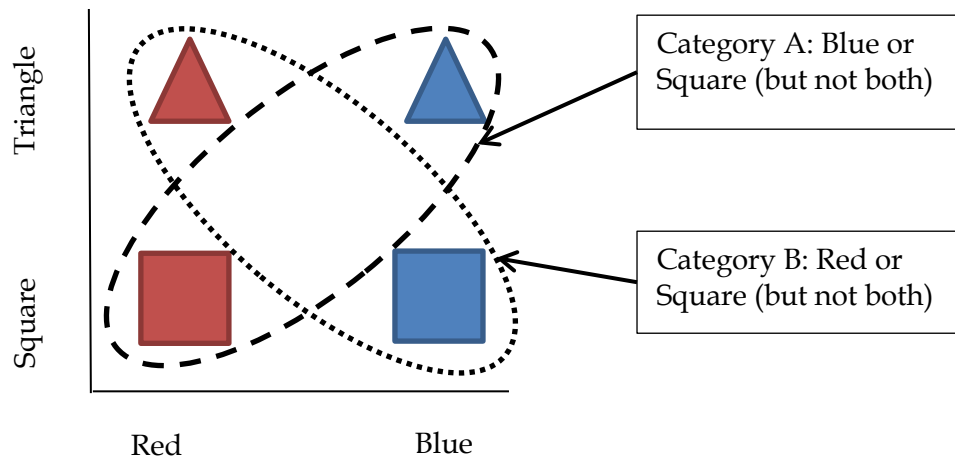
Figure 3: Nonlinearly separable categorisation

In both Figure 2 and Figure 3, in order to make a correct categorisation, it is necessary to consider both the colour and shape of the object. It is impossible to make a decision on the basis of a single dimension. This is known as an unreducible decision. In contrast, if it were possible to decide on the basis of a single dimension, as in Figure 1 where shape is irrelevant, this would be a reducible decision.

The examples discussed above have only two dimensions that contribute to category membership. However, category membership can be determined by an infinite number of categories. Where more than two categories determine category membership, linear separability is established if a hyperplane can be drawn that separates true from false dimensions. The hyperplane has (n-1) dimensions, where n = the number of dimensions that determine category membership (Blair and Homa, 2001). For instance, where category membership is determined by three dimensions, as is the case for some of the problems used in this study, linear separability is established by constructing a three-dimensional graph and drawing a two-dimensional plane that separates true from false answers. For illustrations of this, see Figure 26 or Figure 31 in Appendix A.

NLS categorisation problems have been of interest to the machine learning and artificial intelligence (AI) communities for over 50 years. This interest was sparked by Minsky and Papert's (1972) mathematical proof that two layered 'perceptrons' could not solve NLS problems. This represented a potential boundary on the learning ability of AI.

More recently, research on NLS categorisation has extended to human research. This was prompted by an interest in the extent to which humans and AI shared limits on NLS categorisation. Such a finding may have implications for predictive models of human cognitive performance. Research findings to date suggest that there are constraints on the extent to which humans can learn NLS categorisation, demonstrated through longer time taken to learn categorisation rules, and higher error rates when making categorisation decisions (Ashby et al., 2001; Ell and Ashby, 2006; Maddox et al., 2004; Rehder and Hoffman, 2005; Smith et al., 2011).

It is suggested that the reason NLS categorisation is more difficult to learn than LS categorisation is because people tend to make categorisation decisions on the basis of objects' similarity (Blair and Homa, 2001). Members of LS categories are usually more similar than members of NLS categories, for instance, in Category A in Figure 2 (LS categorisation) two objects are the same colour, and two are the same shape. In contrast, the members of Category A in Figure 3 (NLS categorisation) are not the same shape, or the same colour.

While the difficulties of learning NLS categorisation are clear, the relationship between separability and inference is unclear. While he did not directly compare NLS and LS categorisation, two studies by Yamauchi have demonstrated that both LS (Yamauchi and Markman, 1998) and NLS (Yamauchi et al., 2002) categorisation rules are more difficult to learn through inference than through classification. However, Markman and Ross (2003) suggest that LS categorisation is more easily learned through inference than classification, whereas the reverse is true for NLS categorisation.

1.4   Mental models and the Johnson-Laird effect

We suggest that the difficulties in learning NLS categorisation (Blair and Homa, 2001; Ell and Ashby, 2006; Maddox et al., 2004; Rehder and Hoffman, 2005; Smith et al., 2011) may explain a common finding in psychology, the difficulty in solving complex reasoning problems such as:

> Only one statement about a hand of cards is true:
> (1) There is a King or Ace or both
> (2) There is a Queen or Ace or both
> Which is more likely, King or Ace?
> (Johnson-Laird and Savary, 1996; Johnson-Laird and Savary, 1999).

When asked to solve this problem, the majority of people answer Ace (Johnson-Laird and Savary, 1996; Johnson-Laird and Savary, 1999). However, this answer is incorrect, as it does not take into account that when one statement is true, the other must be false. That is, if Statement 1 is true, and the hand contains a King or an Ace or both, then Statement 2 must be false, and the hand cannot contain a Queen or an Ace or both. Consequently, the hand can never contain an Ace, only a King or a Queen. Therefore the King is more likely to occur than the Ace.

Considering the fact that only one statement can be true at any time is also essential to correctly solving the problem presented in Section 1.1. That problem states that you have the mints, which means that Statement 1 is true. Since only one statement can be true, this means Statement 2 must be false. If you have both the gum and the lollipops, Statement 2 is false (since it explicitly states you cannot have both). However, if you have only one of the gum or the lollipops, this makes Statement 2 true. Hence, the correct answer to the question "Is it possible that you also have either the gum or the lollipops? Could you have both?" is that it is possible to have both, but not possible to have only one.

The difficulty in solving these problems is frequently attributed to the use of mental models when reasoning. According to this explanation, people construct mental models of possible answers when considering the problem (Johnson-Laird and Savary, 1996; Johnson-Laird and Savary, 1999). However, as the complexity of the problem increases, it becomes more difficult to keep track of all possible answers and relevant information. Consequently, people begin to omit information to keep the mental model to a manageable size. In particular, explicitly false information will be omitted from the model. While this keeps the problem within the limits of working memory, it introduces logical errors, as people fail to consider the implications of the false statement.

Under the mental models theory, Johnson Laird and colleagues predict that when people are required to consider false information, such as in the above problem, the use of partial mental models will lead to incorrect answers. In contrast, where people are not required to consider false information, the use of partial mental models will not lead to incorrect answers. These findings have been replicated by Johnson-Laird and other researchers, including a study conducted by DSTO researchers (Sparkes and Huf, 2003).

The DSTO study used versions of Johnson-Laird's problems, modified so they were written in military terminology, e.g.:

> Only one of the following statements about an impending enemy attack is true:
> (1) The enemy will approach from Wade valley or Swain valley or both.
> (2) The enemy will approach from Swain valley and artillery fire will warn of their approach.
> Is it possible for the enemy to come from Swain valley and for artillery fire to warn of their approach?

Participants were six military personnel and six civilians. Sparkes and Huf (2003) found that the military participants were significantly faster to respond than civilian participants, but there was no significant difference between the groups in the number of correct responses.

1.4.1   Linear separability explanation for the Johnson-Laird effect

The problems used by Johnson-Laird and colleagues are categorisation and inference problems, although they do not use this terminology. For instance, in the King and Ace problem discussed above, there are a range of cards in the hand that are logically possible, and a range of cards that are logically impossible. In order to determine whether the King or the Ace is more likely, people must first determine if each card is logically possible or impossible. This is a categorisation decision. If this process is conducted correctly, the logical impossibility of the Ace will be clear, and the correct answer will be achieved.

Some problems Johnson-Laird uses in other studies are inference problems, such as the following:

> Suppose that at least one of the following assertions is true, and possibly both:
> (1) You have the marshmallows.

(2) You have the truffles or the Jolly Ranchers, and possibly both.
Also, suppose you have the marshmallows. What, if anything, follows? Is it possible that you also have either the truffles or Jolly Ranchers? Could you have both?
(Khemlani and Johnson-Laird, 2009)

This is an inference problem because the category membership is known (logically possible combinations), as is one of the characteristics used to define category membership (marshmallows present). Solving the problem requires identification of other characteristics (truffles and Jolly Ranchers present or absent)[5].

If Johnson-Laird's problems can be considered categorisation and inference problems, then the linear separability of the problems may affect the extent to which they are easily solved. We have conducted analysis that supports this. As discussed in detail in Appendix A, we analysed 14 problems from six of Johnson-Laird's studies (Goodwin and Johnson-Laird, 2010; Goodwin and Johnson-Laird, 2011; Johnson-Laird and Savary, 1996; Johnson-Laird and Savary, 1999; Khemlani and Johnson-Laird, 2009; Santamaría and Johnson-Laird, 2000). We determined whether each problems was LS or NLS, and examined the percentage of participants in the original studies who correctly answered LS and NLS problems.

While full analysis and worked examples are contained in Appendix A, Figure 4 shows a summary of the percentage of participants in each study who correctly solved LS and NLS problems. Each column in the figure refers to a single problem used in a specific study. The figure shows that LS problems were solved by the majority of participants, with the percentage of correct answers ranging from 62-100%. In contrast, with one exception, NLS problems were not solved correctly by the majority of participants. Omitting the single NLS problem that was solved correctly by 100% of participants, the percentage of correct answers for the remaining NLS problems ranged from 0-48%.

Based on this analysis, we believe that the linear separability explanation may be a plausible explanation for the Johnson-Laird effect. However, Johnson-Laird does not appear to have considered this explanation. In addition, it is not possible to simply re-analyse Johnson-Laird's problems to test the linear separability explanation. As discussed in Appendix A, there are possible confounds from factors such as the number of terms used in the problems, the extent to which all terms need to be considered to solve the problem, and the level of clarity and concreteness of the problem. Hence, the current study was developed to test the linear separability explanation.

---

[5] If you have the marshmallows, then Statement 1 is true. The problem states that either or both statements can be true. Hence, the possible outcomes are that Statement 2 is false, and you have no additional confectionary, or that Statement 2 is true, and you have either or both of the truffles and the Jolly Ranchers. For further analysis of this problem, see Problem 10 in Appendix A.
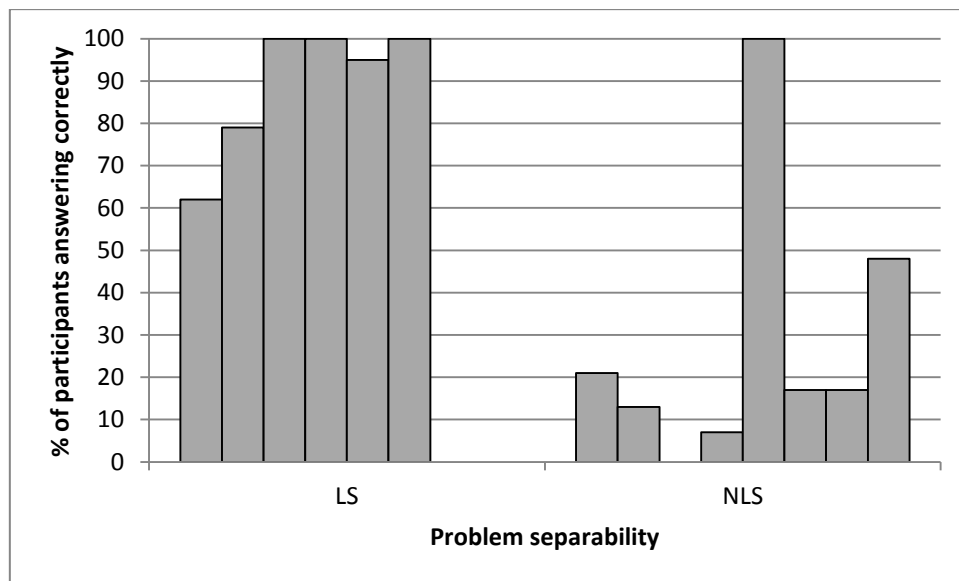
Figure 4: Percentage of participants correctly solving LS and NLS problems based on reanalysis of Johnson-Laird's data (see Appendix A)

1.5 The current study

Johnson-Laird suggests that problems are more complex to solve when they require falsification of the mental model. We suggest that the complexity arises because the problems are NLS. The current study was designed to test this explanation. The experimental proposal was derived from Johnson-Laird's work, but with modifications as follows.

First, in Johnson-Laird's studies, participants are given the rule (i.e. "Only one of the following statements is true…"), and a single case to test against the rule ("suppose you have X. Can you have Y?"). In this study, participants were required to learn the rule, through repeated presentations of all possible combinations of variables. Participants were also required to make multiple inference judgements.

Second, this study did not use written problems. Some researchers (Barrouillet and Lecas, 2000) have suggested that Johnson-Laird's findings can be attributed to participants misreading or misunderstanding the questions[6]. As a problem, this study used a light

---

[6] For instance, consider the statement '*Suppose that you are playing cards and that you get two cards. You know that if the first card is a king, then the second card is an ace, or else if the first card is not a king, then the second card is an ace*'. Using the same logic as Johnson-Laird's problem on p7, it should be clear that this problem contains two statements '*You have a king and an ace*' and '*You don't have a king and you have an ace*'. If only one of these statements can be true, it is impossible for an ace to be in the hand. However, Barrouillet and Lecas (2000) suggest that people interpret the statement as '*you have a king, or you don't have a king, and you have an ace*', which leads to the incorrect conclusion that the ace is logically possible. The misunderstanding arises because of the way people interpret 'or else' in the above statement.

switch, controlled by three different shaped light switches. No written material was used to describe the problems, only particular combinations of switches, and the state of the light. This was intended to remove any possibility that misunderstanding or misinterpretation of the problems contributed to difficulties in solving them.

The paradigm used in this study was a light, controlled by three light switches. Whether the light was switched on or off was determined by a LS or NLS function. Participants first learned the rule through repeated categorisation decisions, and then were tested on their ability to make inferences.

This study had two hypotheses. The first was that NLS categorisation would be more difficult to learn than LS. This would be demonstrated in the Categorisation phase through:
- Lower rates of correct responses across all trials,
- More trials to reach 100%,
- Fewer trials with a score of 100%, and
- Slower response times across all trials.

The second hypothesis was that NLS functions would be more difficult to comprehend than LS functions. This would be demonstrated through lower rates of correct responses and slower response times in Inference.

2. Method

2.1 Participants

Participants were 32 military and civilian personnel from an Australian Army regiment and DSTO. Ages ranged from 21 to 50 years, with an average age of 32 years old.

2.2 Materials

One LS function to serve as a practise items, and five LS and five NLS functions to serve as test functions were generated. Each function contained three variables, each with two values, true or false. This means that for each function, there were eight (or 2^3) possible combinations of variables. The functions were generated and selected according to the following criteria. First, each function had three instances where the light was switched on, and five where the light was switched off. Second, the functions were irreducible, meaning that in all cases, it was necessary to consider all three variables to solve the function. A full list of the functions is contained in Appendix A.

A word search puzzle downloaded from http://www.Printable-Puzzles.com was used as a filler task after participants had completed testing on the first function. The puzzle was

intended to reduce carryover effects and interference between the first and second function.

2.3   Design and Procedure

The study employed a within-subjects design, testing categorisation and inference of LS and NLS functions. On arrival, participants were given a brief on the study, and gave informed consent to participate. Participants then read through task instructions (see Appendix C), and completed a short practise of Categorisation and Inference judgements. Once participants had completed the practise, and were confident they understood all instructions, the experiment proper commenced.

Each participant was tested on one LS and NLS function, randomly selected from the pool of five functions. Half the participants were tested on the LS function first, and half were tested on the NLS function first. For each function, there were two components, Categorisation, and Inference. In Categorisation, participants were shown a combination of the three light switches, such as in Figure 5, and had to judge if the light was on or off. Immediate onscreen feedback (CORRECT or INCORRECT) was provided once a response was made. There were eight possible combinations of switches[7] ("one block"). These eight combinations were repeated in every Categorisation block, allowing measurement of participants' learning across the duration of the experiment.

Following eight blocks of Categorisation (64 individual on/off judgements), participants were presented with seven to nine Inference questions[8]. In each, participants were shown a combination of one or two shapes and the light state (on or off), as in Figure 6. This combination of light switches and on or off state was controlled by the same function participants had just learned. Participants were asked what could be deduced about another shape. The response options were: shaded, unshaded, either, or don't know (to discourage guessing). Each Inference question was unique (ie, seen by participants only once through the experiment). This sequence of Categorisation followed by Inference occurred five times for each function, as summarised in Figure 7.

After completing categorisation and inference for one function, participants spent five minutes performing a word search puzzle as a filler task. Following this, the categorisation and inference procedure was repeated for the second function.

---

[7] There were three switches, each with two possible states – on or off. Therefore, the total number of possible combinations of the switches was 2^3, or eight.

[8] The decision to vary the number of Inference questions in each trial pre-dates the first author's involvement in this project. No documentation has been found to explain this decision. It is possible this was done to avoid predictability or to reduce the likelihood that participants could arrive at the correct answer by guessing or making predictions based on the number of previous questions in the trial.

Figure 5: Categorisation decision
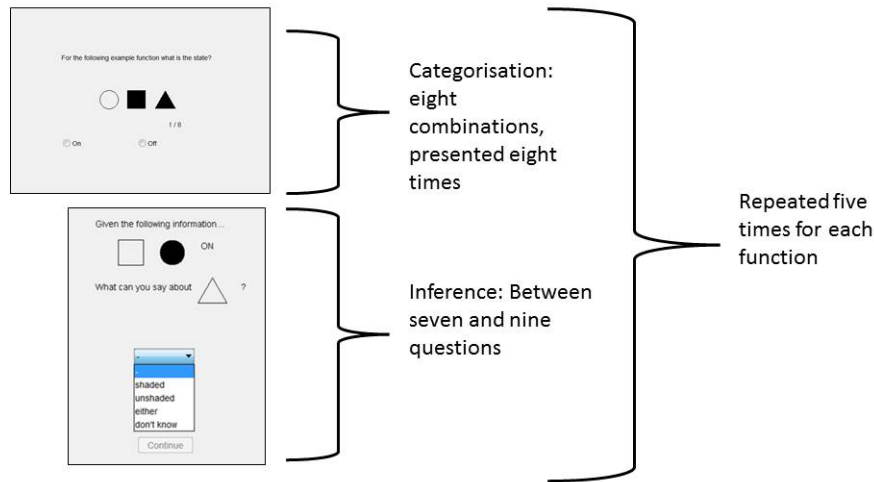

Figure 6: Inference decision

Figure 7: Experiment structure

At the end of the study, participants were asked to fill out a short survey (see Appendix C for a copy) asking:

- which function they found easiest to solve,
- how they solved the function, and
- their confidence in their answers.

This survey was to examine:

- if NLS functions were perceived to be more difficult to solve,
- whether participants were attempting to derive the function, were memorising the correct answers, or using another strategy, and
- if there was any relationship between confidence and accuracy.

Participants were tested in groups. They were instructed to complete the experiment at their own pace, and most participants took between 60-80 minutes. The study received ethics approval in accordance with DSTO's procedures (protocol number LOD 01/12), and was conducted in accordance with research ethics principles (NHMRC, 2007).

3. Results

Unless otherwise indicated, all results are reported to two decimal places. Exact probability values for statistical tests are reported to three decimal places, except where p < .001. The symbol $p\eta^2$ refers to partial eta squared, a measure of effect size.

3.1 Categorisation

Data collected during Categorisation comprised the number of correct responses, and the time taken to respond. As noted in Section 2.3, Categorisation involved five groups of 64 individual on/off judgements, interspersed with Inference judgements. As there were 320 Categorisation judgements for each function, for ease of analysis, they have been grouped into ten trials, each containing 32 individual on/off judgements.

For each of these ten Categorisation trials, the proportion of correct answers was calculated (ranging from zero to one). The average scores for LS and NLS functions are shown in Figure 8. While the minimum possible score was zero, the axis has been truncated in order to show the trend more clearly. The error bars in this and subsequent figures represent the Standard Error of the Mean. When interpreting the Categorisation figures, recall that Categorisation and Inference sequences alternated. Every second Categorisation trial was followed by seven to nine Inference questions (see Figure 7).

In Figure 8, two clear trends are apparent. First, there is a steady improvement in performance across trials, suggesting that learning is taking place, and second, scores are generally higher for LS functions than for NLS. Statistical tests indicated that the improvement across trials was statistically significant, but that the difference in scores between LS and NLS functions was not. Results from a 2 x 10 repeated measures ANOVA, testing the effect of Function Type (NLS, LS) and Trial (1-10) showed that the only significant effect was Trial, $F (9, 270) = 45.12$, $p < .001$ ($p\eta2 = .60$). In addition, paired samples t-tests comparing average scores for LS and NLS functions in each trial (e.g. NLS Trial 1 vs. LS Trial 1) showed that none of the differences were statistically significant.
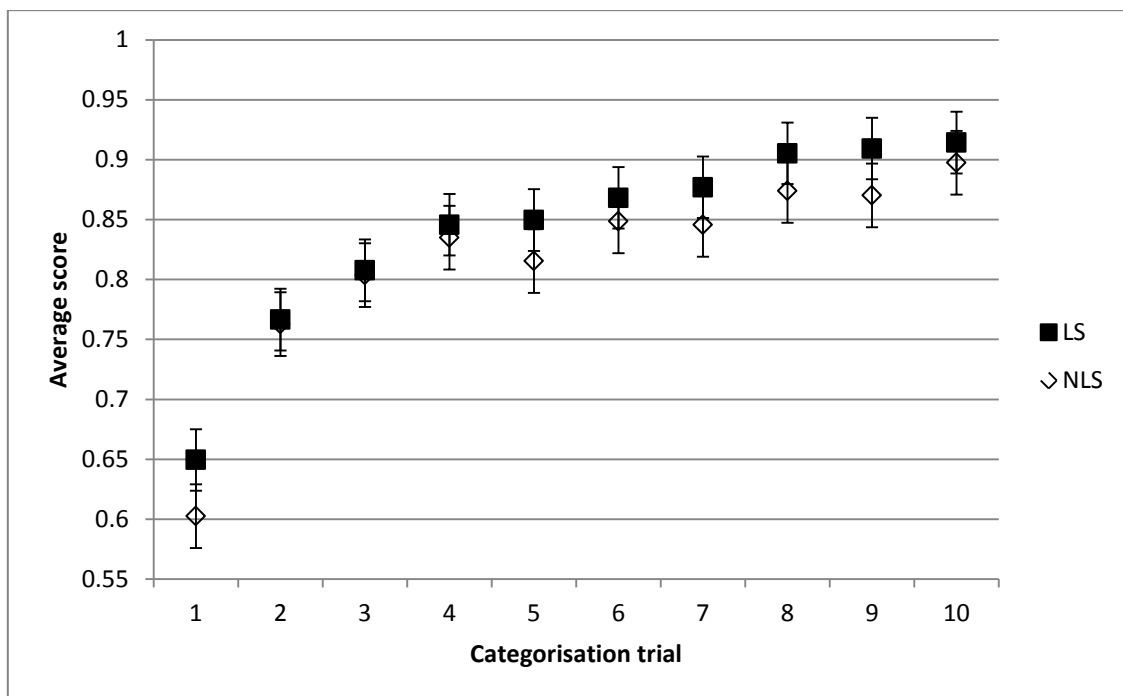
Figure 8: Average score by trial for LS and NLS functions
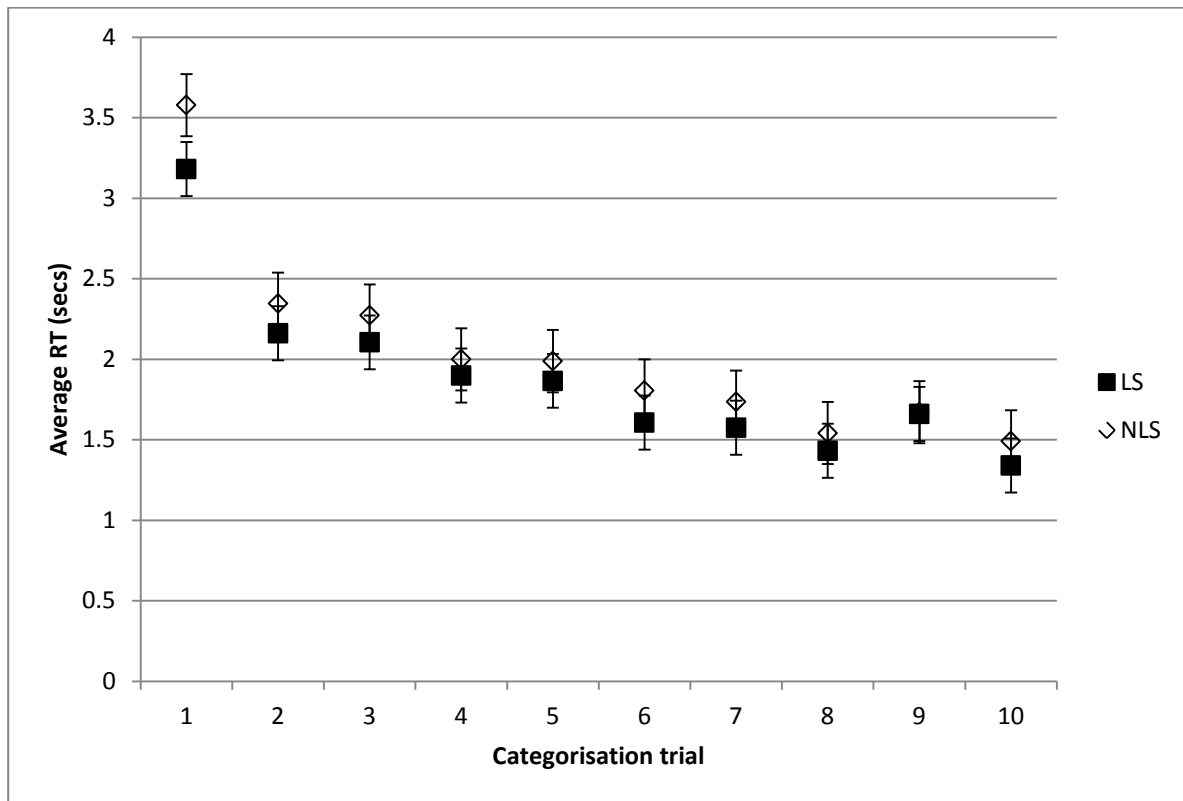
Figure 9: Average response time by trial for LS and NLS functions

The average response time for each trial is shown in Figure 9. On average, response times decreased across trials, with LS functions responded to faster than NLS functions. Results from a 2 x 10 repeated measures ANOVA, testing the effect of Function Type (NLS, LS) and Trial (1-10) showed that the only significant effect was Trial, F (9, 270) = 26.56, $p < .001$ (pη2 = .47). In addition, paired samples t-tests comparing average scores for LS and NLS functions in each trial (e.g. NLS Trial 1 vs. LS Trial 1) showed that none of the differences were statistically significant.

To examine categorisation patterns in more detail, the number of trials taken to reach a score of 100% was calculated. The majority of participants recorded at least one trial with a perfect score (27/32 for LS functions and 28/32 for NLS functions). Figure 10 shows the average number of trials required to obtain a score of 100%. While the average number of trials taken to record a score of 100% was slightly lower for LS functions, this difference was not statistically significant.
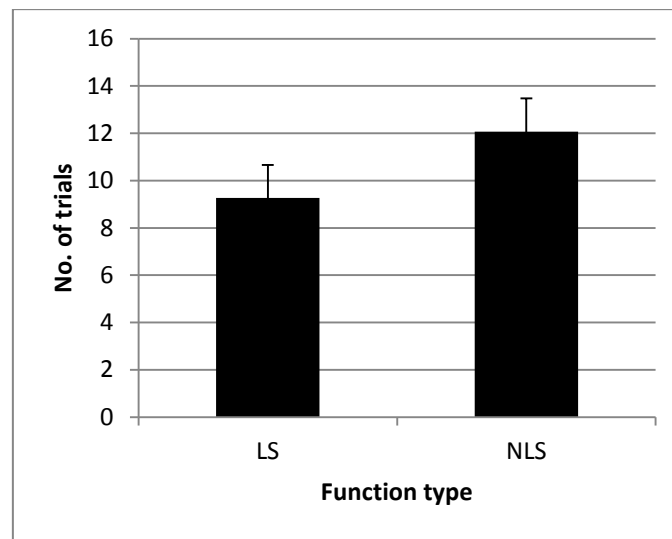
Figure 10: Average number of trials required to reach 100% score for LS and NLS functions

In addition, the average number of trials where a participant obtained a score of 100% was calculated. These results are in Figure 11, and show that on average, 100% scores were obtained more often for LS functions than for NLS. However, this difference was not statistically significant.

Figure 11: Average number of trials where a 100% score was obtained for LS and NLS functions

3.1.1   Analysis by different performance levels

Overall these data suggest a tendency for LS functions to be learned more quickly and accurately than NLS functions, but this was not statistically significant. One possible explanation for these findings was that some participants found both the LS and NLS

functions too difficult to learn. In order to examine this in more detail, the data were divided on the basis of number of times they scored 100% in LS Categorisation judgements into three groups: Top Performers ($n = 11$), Middle Performers ($n = 11$), and Bottom Performers ($n = 10$).

Dividing participants into groups could have been done on the basis of a number of different measures, e.g., the number of trials taken to reach 100% for either LS or NLS functions, average score across all categorisation trials, or results in Inference. The measure chosen for categorisation was arbitrary. However, it was strongly correlated ($r = .94$) with average score across all LS Categorisation Trials, with a large to very large correlation with average score across all NLS Categorisation Trials and the number of times 100% was scored for NLS Categorisation Trials ($r = .61$ for both correlations, effect sizes description from Hopkins, 2002).

Although the majority of participants recorded more 100% scores for LS than NLS functions, for six participants, this trend was reversed. That is, they scored more 100% trials for NLS functions than LS. These six participants all solved a LS function first, followed by a NLS function. Hence, their higher score for NLS functions may indicate practice effects. Three of these participants were in the Middle Performers Group, and the remaining three were in the Bottom Performers Group. The implications of this are discussed later in the report (see Figure 13 and associated discussion).

The Categorisation data for the three groups are contained in Figure 12. It is clear from the figure that the learning patterns for Bottom Performers differ markedly from those for Top and Middle Performers; while the latter two groups' average approaches ceiling, the average performance of the Bottom Performers does not exceed 80% on the last trial.

A 3 x 2 x 10 mixed ANOVA was conducted on these data, examining the effects of group (Top, Middle, Bottom), function type (LS, NLS), and Trial (1-10). This showed that the following main effects and interactions were statistically significant:
- Group, $F (2, 28) = 27.69$, $p < .001$ ($p\eta2 = .66$)
- Trial, $F (9, 252) = 57.51$, $p < .001$, ($p\eta2 = .67$)
- Trial x Group, $F (9, 252) = 5.27$, $p < .001$ ($p\eta2 = .27$)
- Function x Trial x Group, $F (18, 252) = 1.67$, $p = .045$ ($p\eta2 = .11$).

The final significant interaction, between function, trial, and group, suggests that that was a significant effect of linear separability for at least some of the groups on some of the trials. In order to further explore this, a series of paired samples t-tests was conducted, examining the difference between average scores for LS and NLS functions for each group in each trial. That is, LS vs. NLS Bottom group trial 1, LS vs. NLS Middle group Trial 1, etc. Results from these tests indicated that in the Middle group, there were significant differences between average scores for NLS and LS functions in 2 trials:
- Trial 7, $t (10) = 2.66$, $p = .024$ (Cohen's $d = -0.30$)
- Trial 9, $t (9) = 2.45$, $p = .037$ (Cohen's $d = -1.23$).

These were the only significant comparisons in all three groups.

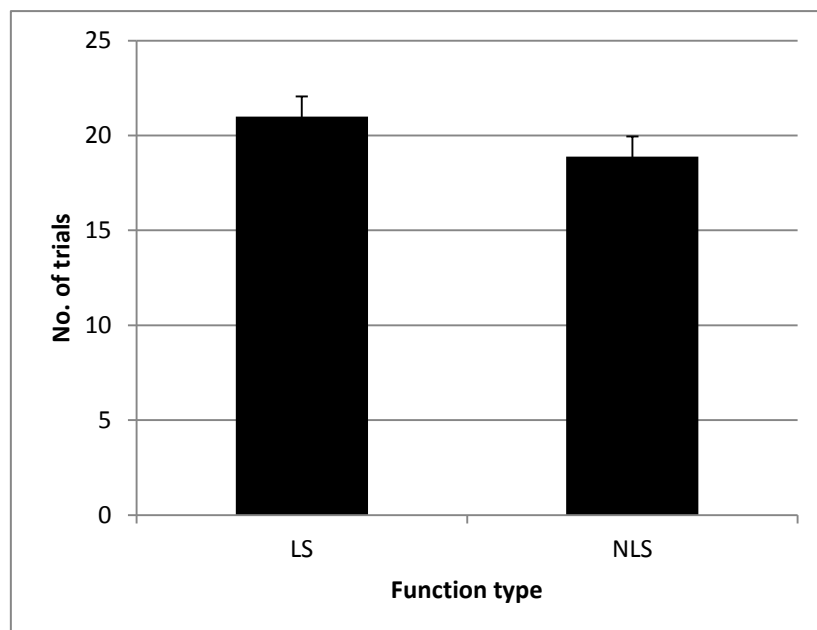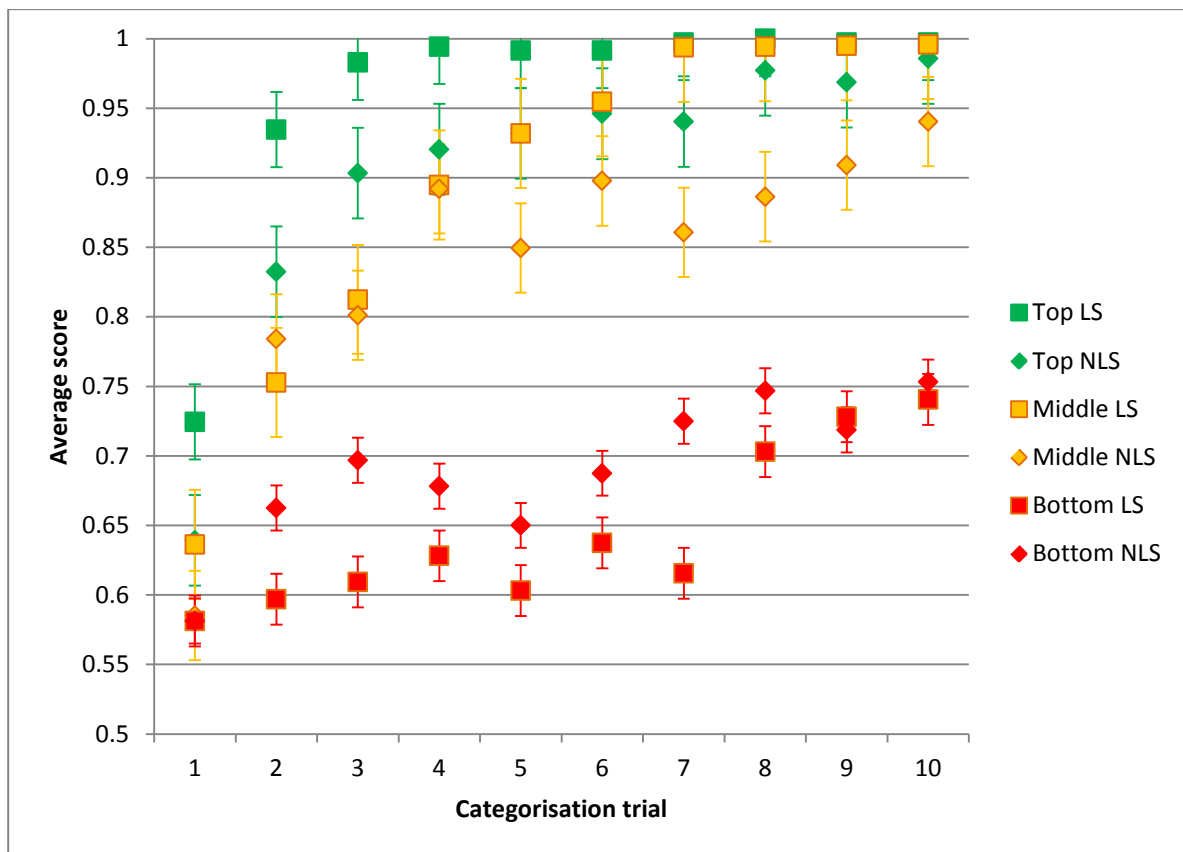*Figure 12: Average score by trial for LS and NLS functions by group*
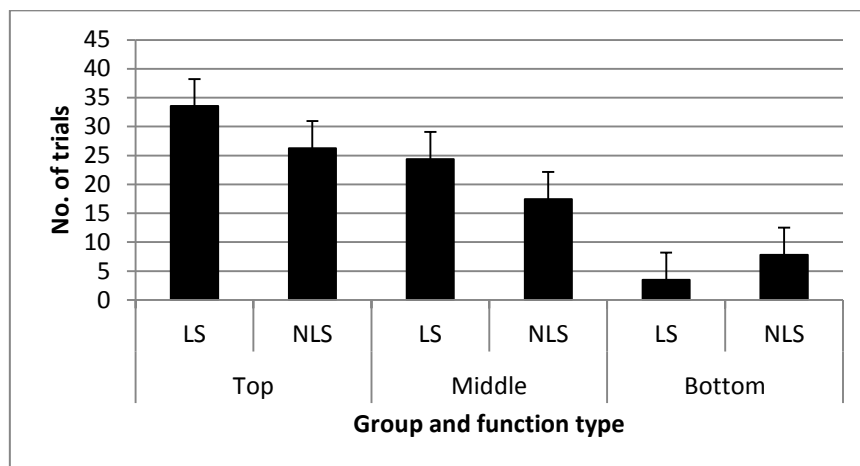

Figure 13: Average number of trials where a 100% score was obtained for LS and NLS functions by groups

Results for the number of trials where a 100% score was obtained are shown in Figure 13. It is clear from the figure that for the Top and Middle Performers groups, there were more 100% scores obtained in LS trials than NLS. This difference was statistically significant for

the Top group, t (10) = 2.17, p = .02 (Cohen's d = 1.28). This difference approached statistical significance for the Middle group, t (10) = 1.90, p = .09 (Cohen's d = .82).

In the Bottom Performers group, more 100% scores were recorded for NLS functions than for LS, although this difference was non-significant. As discussed earlier in this section, three participants in the Bottom Performers Group recorded more 100% scores for NLS functions than LS functions. This appears to be a practice effect, as all three participants solved LS functions first, followed by NLS functions.

Results for the number of trials taken to reach a score of 100% are shown in Figure 14. As the figure shows, in the Top group, participants reached criterion in an average of five trials for LS functions, and ten trials for NLS functions. However, this difference was not statistically significant, t (10) = 1.59, p = .14. Closer examination of the Top group data revealed that the average trials to criterion for the NLS functions were skewed by one participant who took 38 trials to reach criterion. When this participant's results were removed, the difference between NLS and LS functions approached levels of significance, t (9) = 2.03, p = .07 (Cohen's d = .80).

The differences between trials to criterion for the Middle and Bottom groups were not significant.

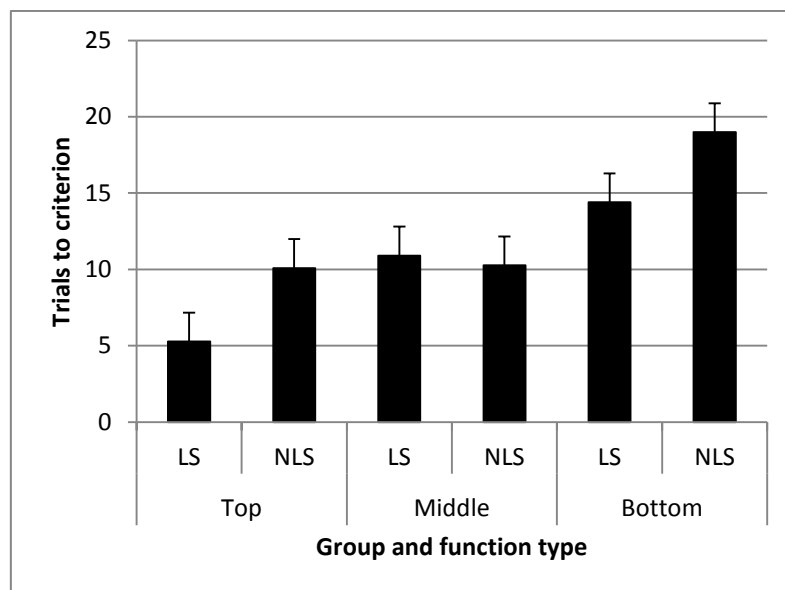*Figure 14: Trials to criterion by group and function type*

The average response time for each trial by Group is shown in Figure 15. A 3 x 2 x 10 mixed ANOVA was conducted on these data, examining the effects of Group (Top, Middle, Bottom), Function type (LS, NLS), and Trial (1-10). This showed that the only significant effect was Trial, F (9, 252) = 25.71, p < .001 (pη2 = .48).
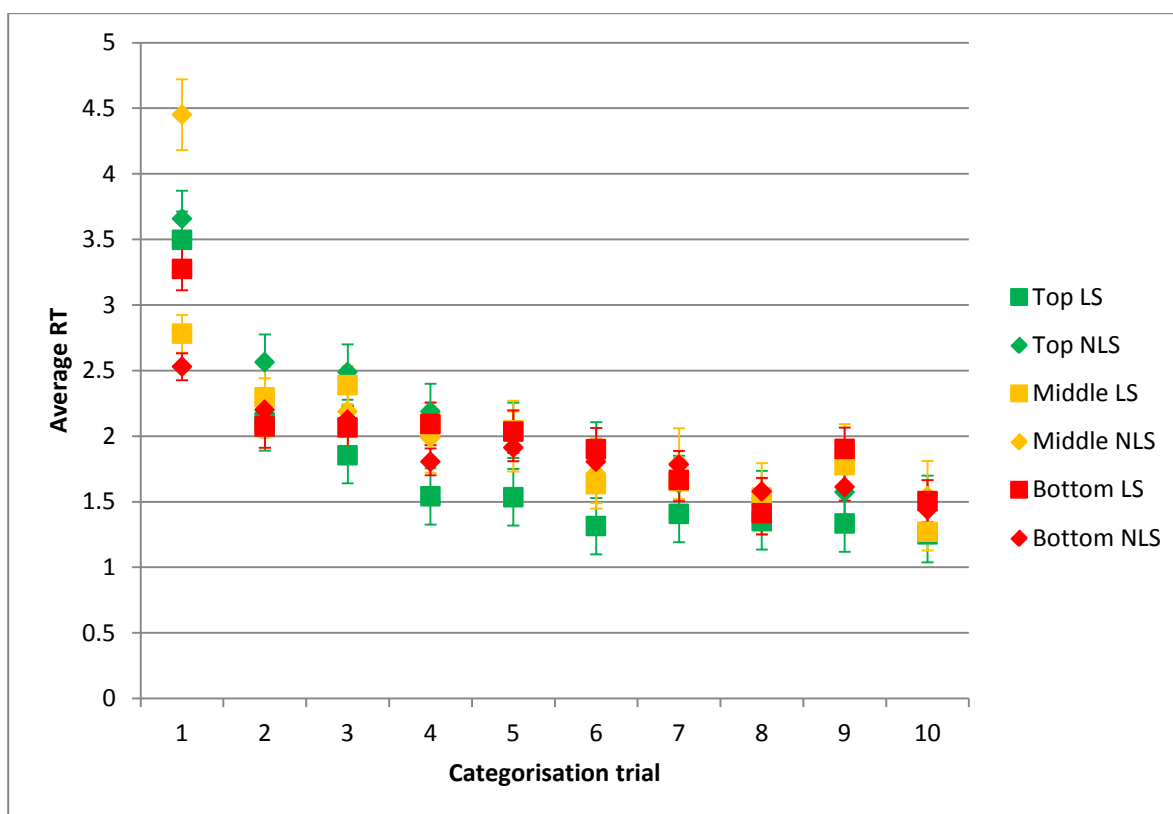
Figure 15: Average RT by trial for LS and NLS functions by group

3.2   Inference

Data collected during the inference trials comprised the number of correct responses, and the response time. As discussed in Section 2.3 and Figure 7, there were five groups of seven to nine inference questions following a series of Comprehension questions. For ease of analysis, each group of inference questions was considered the base unit of analysis. These groups were designated 'trials' to keep consistency with the naming conventions used for Comprehension.

The average score for each inference trial for LS and NLS functions is shown in Figure 16. Average scores are higher for LS functions for two inference trials, higher for NLS in two inference trials, and approximately equal for the remaining inference trial. A 2 x 5 repeated ANOVA testing the effects of Function (NLS, LS) and Trial (1-5) showed that the only significant effect was Trial, $F_{(4, 120)} = 4.33$, $p = .003$ (pη2 $\eta p^2 = .13$). A series of paired samples t-test comparing the difference between LS and NLS scores for each trial showed that the difference was significant for Trial 4 only, $t_{(31)} = 2.32$, $p = .03$ (Cohen's $d = 0.31$). Note that this difference is in the opposite direction to that predicted by the hypothesis. That is, performance was significantly *better* for NLS than LS.
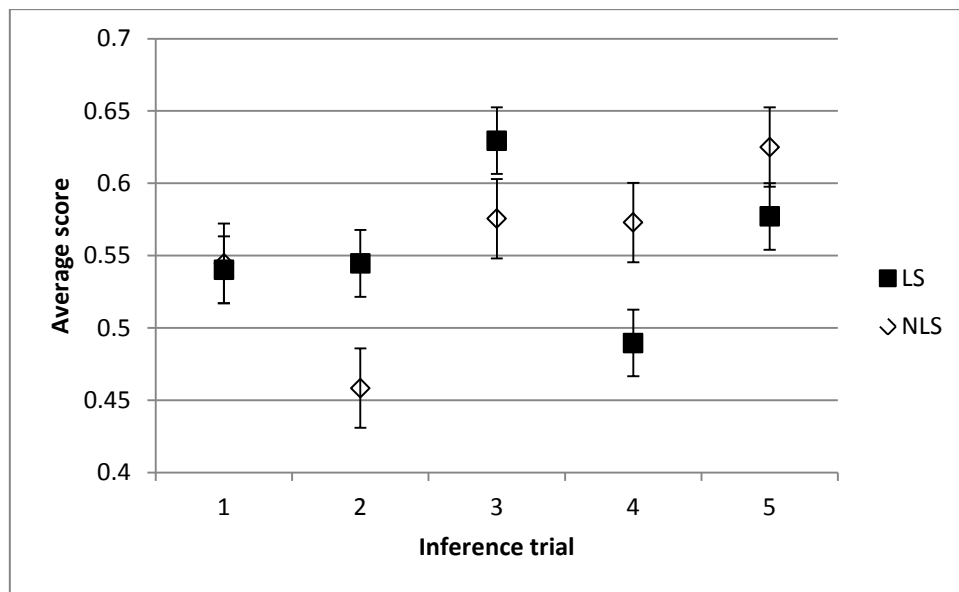
Figure 16: Average score by inference trial for LS and NLS functions

The average response time for each inference trial for LS and NLS functions is shown in Figure 17. A 2 x 5 ANOVA testing the effects of Function (NLS, LS) and Trial (1-5) showed that the only significant effect or interaction was Trial, $F(4, 120) = 7.23$, $p < .001$ (pη2 = .19). A series of paired samples t-tests comparing the difference between LS and NLS reaction times for each trial showed that the only significant difference occurred in Trial 1, $t(31) = 2.262$, $p = .031$ (Cohen's $d = .41$).
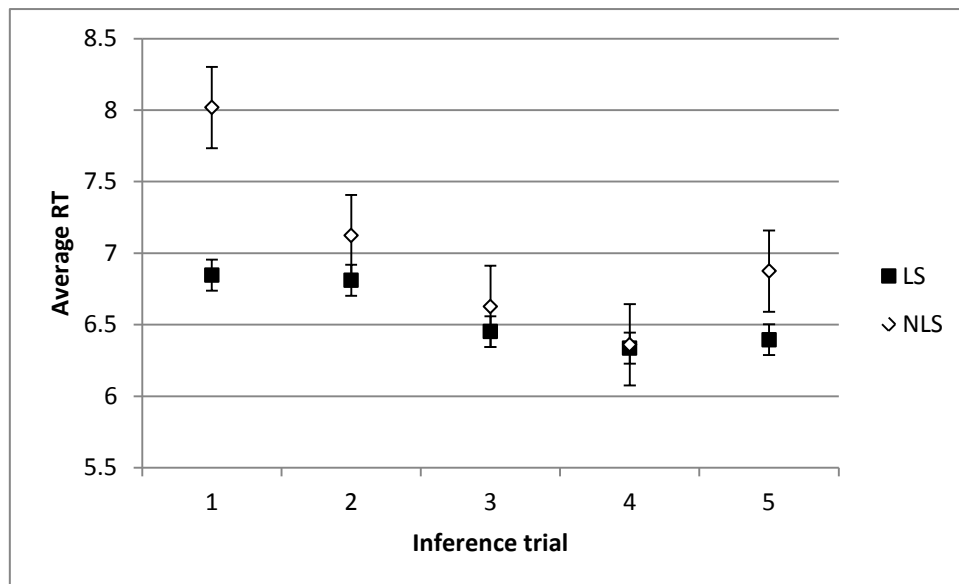
Figure 17: Average response time by Inference trial

As with the data from Categorisation, the proportion of correct responses in Inference were divided into Top, Middle and Bottom performers. These results are shown in Figure 18. It is clear from the figure that there is a strong trend for the Top group to score higher than the Middle and Bottom groups. In addition, there is a trend, particularly in the

Top group, for scores to be higher for LS functions than NLS. However, there are a number of instances, particularly in the Bottom group, where scores are *higher* for NLS than LS functions.

A 3 x 2 x 5 mixed ANOVA was conducted on these data, examining the effects of group (Top, Middle, Bottom), function type (LS, NLS), and Trial (1-5). This showed that the following main effects and interactions were statistically significant:

- Trial, $F (4,112) = 4.376$, $p = .003$ (pη2 = .14)
- Function x Group, $F (2, 28) = 3.596$, $p = .041$ (pη2 = .20)
- Function x Trial x Group, $F (8, 112) = 2.847$, $p < .001$ (pη2 = .17).

Paired samples t-tests conducted on the Top performing group indicated that the difference between LS and NLS functions was statistically significant for Trial 2 only, $t (10) = 3.16$, $p = .01$ (Cohen's $d = 1.30$). The difference between Trials 1 and 4 approached levels of statistical significance, with p values of .09 and .07 respectively. No comparisons were significant for the Middle and Bottom groups.
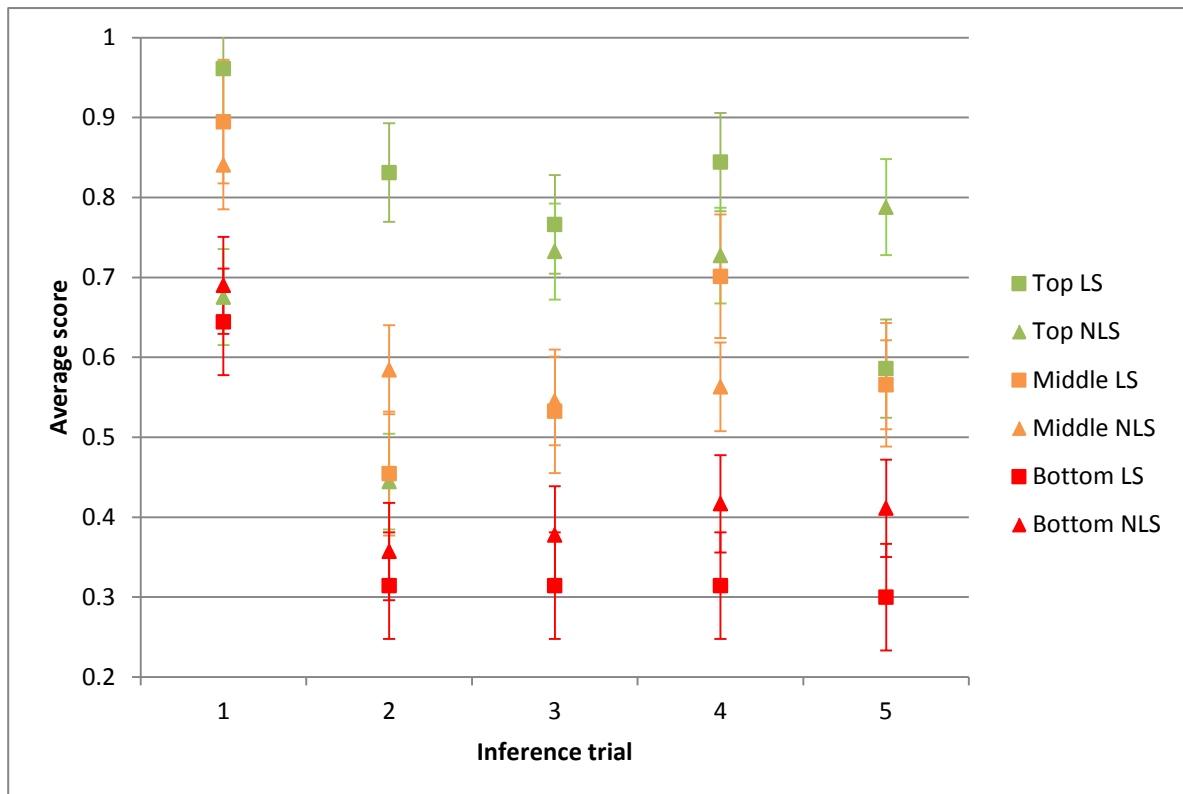

Figure 18: Average score by inference trial by group and function type

The response time data were further analysed on the basis of groups. These data are shown in Figure 19. It is clear from the figure that the poorest performing participants, the Bottom Performers group, had average response times considerably faster than the other two groups.
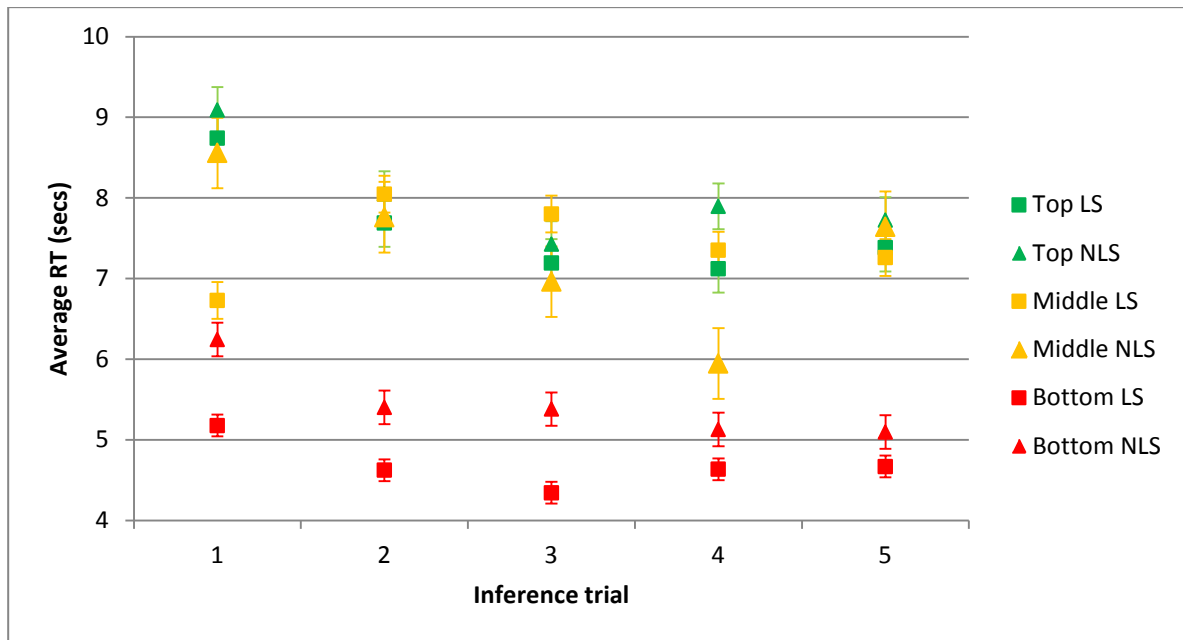
Figure 19: Average response time by Inference trial, group, and function type

A 3 x 2 x 5 mixed ANOVA was conducted on these data, examining the effects of group (Top, Middle, Bottom), function type (LS, NLS), and Trial (1-5). This showed that the only significant effects were Trial, $F_{(4, 112)} = 7.224$, $p < .001$ ($p\eta2 = .21$), and Group $F_{(2, 28)} = 4.849$, $p = .016$ ($p\eta2 = .26$).

3.2.1 Analysis of "Don't Know" responses

As indicated in Section 2.3, "Don't Know" was one of the response choices when making inference judgements. Across the experiment, seven participants – three in the Top performing group, one in the Middle performing group, and five in the Bottom performing group[9], gave this response at least once during Inference judgements.

A frequency distribution of the "Don't Know" responses by function type and group is shown in Figure 20. It is clear from the figure that Top and Middle performing participants were more likely to answer "don't know" in response to NLS functions, whereas Bottom performing participants were more likely to answer "Don't Know" in response to LS functions. A 2 x 3 Chi-square analysis indicated that the distribution of responses was statistically significant, $X^2 (2) = 22.399$, $p < .001$. However, it is possible that the results of this test are skewed due to the small number of participants recording "Don't Know"

---

[9] This included one participant who answered "don't know" to all inference questions. While it was initially difficult to determine if this participant genuinely did not know the correct answer, or was disengaged from the study, examination of their categorisation trials showed a learning curve consistent with other participants. That is, after starting at approximately chance levels, performance slowly increased to close to ceiling. On this basis, it's concluded that the participant was genuinely engaged with the experiment but simply found Inference too difficult. Hence, their data were included in analysis.

responses, the single participant who failed to answer a single inference question correctly, and the number of observed frequencies fewer than five.
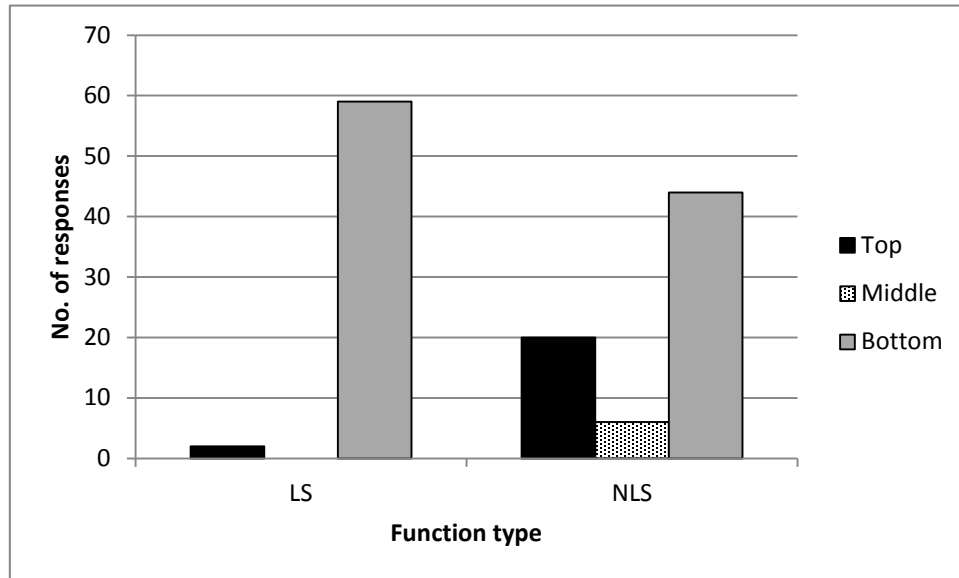
Figure 20: Distribution of "Don't Know" responses by function and group

3.3   Survey data

As described in Section 2.3, at the conclusion of the study, participants completed a short survey. Due to some missing responses, the results were available for only 30 participants.

The first question asked participants which problem they found easiest to solve, the first problem, the second problem, or both. The purpose of this question was to identify if participants perceived that the LS problems were easier to solve than the NLS.
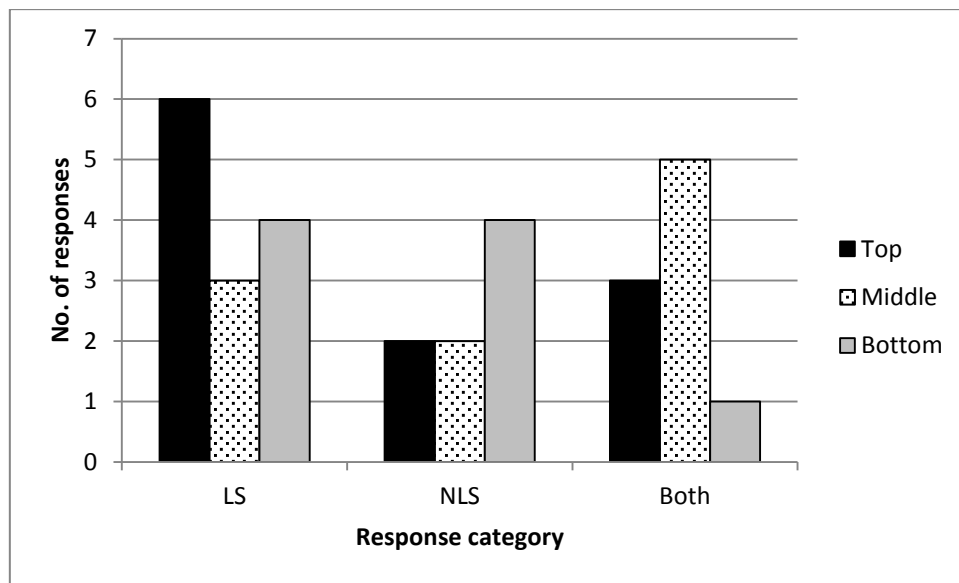
Figure 21: Easier problem to solve by group

Figure 21 shows the results, divided into Top, Middle, and Bottom performing groups. It is clear that in the Top performing group, there was a strong trend for the LS problems to be rated easier to solve. However, in the Middle and Bottom performing groups, responses were more evenly distributed across the three categories. A 3 x 3 Chi-squared analysis showed that this distribution of results was not statistically significant, X^2 (4) = 4.693, p = .32. However, this analysis may have been skewed by the small number of observed frequencies in some cells.

The second question asked participants what strategies they used to solve the problems. The four options were, memorising the correct answers, deducing the underlying rule, guessing, and other. The frequency of responses is shown in Figure 22. It is clear that the majority of participants used deduction, and that participants from the Bottom group were more likely than the Top and Middle groups to use other strategies, such as guessing. A 3 x 3 Chi-squared analysis conducted on these data showed that the distribution of results was not statistically significant, X^2 (6) = 9.16, p = 0.16.
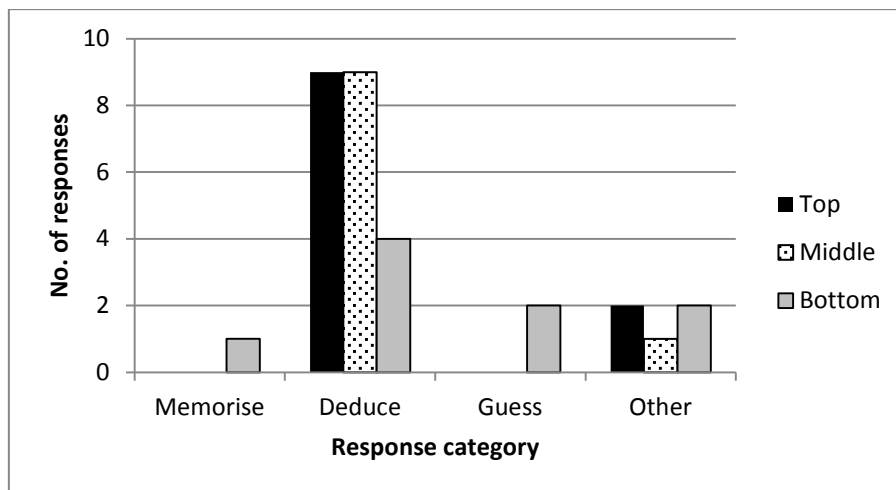
Figure 22: Response strategies by group

The third question asked participants to rate their confidence in solving the problems. The three options were, Not At All Confident, Moderately Confident, and Very Confident. The frequency of responses by group are shown in Figure 23. It is clear that the majority of participants rated themselves as Moderately or Very Confident that they had answered correctly. Only a small number of participants from the Bottom performing group were Not at all Confident. A 3 x 3 Chi-squared analysis showed that the pattern of results were statistically significant, X^2 (4) = 12.76, p = 0.01. However, this should be treated with some caution as the small number of cases in some cells may make the Chi-square unreliable.

Figure 23: Confidence ratings by group

4. Conclusion

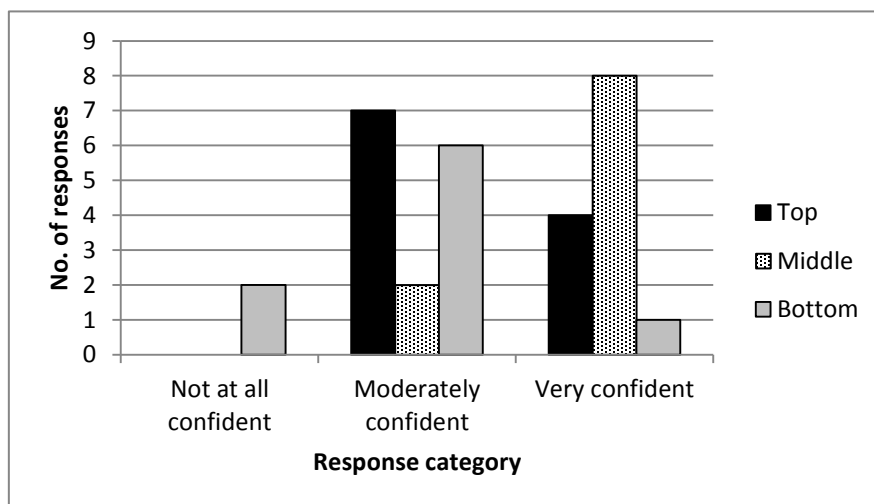The aim of this study was to examine a linear separability explanation for Johnson-Laird's findings (Johnson-Laird and Savary, 1996; Johnson-Laird and Savary, 1999). It was hypothesised firstly, that it would be more difficult to make categorisation decisions for NLS functions than LS, and secondly, that it would be more difficult to make inference judgements for NLS functions than LS. The first hypothesis was consistent with previous research on category learning (Ashby and Maddox, 2005; Ashby et al., 2001; Blair and Homa, 2001; Ell and Ashby, 2006; Rehder and Hoffman, 2005; Smith et al., 2011), while the second hypothesis was intended to demonstrate the plausibility of the linear separability explanation. Results from the study provided only limited support for each hypothesis.

4.1   Categorisation

The first hypothesis was that there would be significant differences between LS and NLS categories in average scores and response times during categorisation. While there were no overall differences, when the participants were divided into groups, some interesting trends became apparent.

The three groups – Top, Middle, and Bottom Performers – each recorded different patterns of results. Average scores in the Top Performers' group quickly approached ceiling for both NLS and LS functions. It appears this group was able to learn both functions equally well. Average scores in the Bottom Performers' group also did not differ for LS and NLS functions. However, in this group, performance reached 75%; this was poor compared to the other two groups. While not low enough to be considered a floor effect, this implies the participants were unable to correctly learn the categorisation rules.

In contrast to the Top and Bottom Performers groups, average scores for the Middle Performers' group showed significant differences between LS and NLS functions for two trials. The learning curves for this group (see Figure 12) suggest that categorisation of the LS functions approached ceiling, while categorisation for NLS was markedly poorer.

When considering these results, it is noteworthy that the response times did not differ between function types, or between groups. In addition, when participants were asked in the post-experiment survey what type of strategy they used to solve the problems, all groups showed a strong preference for attempting to deduce the rule, rather than memorising the correct responses. This suggests that the superior performance of the Top Performers compared to the Bottom Performers did not come about because they took more time to think about the correct answer, or because they used a more effective strategy. Similarly, the failure of the Bottom Performers group to differentiate between LS and NLS functions did not occur because they spent different amounts of time thinking about how to respond, or because they used a less effective strategy. It appears some people are simply better than others at categorisation. This is consistent with results obtained by the second and third authors and colleagues (Temby et al., 2005). In their study, testing marksmanship performance, approximately one third of the participants

either were not engaged in the task, or lacked the capability to perform the task. Future research may examine the impact of adopting some form of screening process to remove participants who are apparently unable to perform the task to high levels of performance; this may help provide better differentiation between LS and NLS functions.

The post-experiment survey also suggested that participants were not able to accurately assess their performance. Nearly all participants reported that they were "Moderately Confident" or "Very Confident" that they solved the problem correctly, and there was no clear indication that the LS problems were perceived as easier to solve (except in the Top Performers group, and as discussed previously, this group's performance did not, overall, differ significantly between LS and NLS problems).

The results from the categorisation phase are not wholly consistent with previous research, where significant differences between LS and NLS categorisation have been consistently demonstrated (Ashby et al., 2001; Ell and Ashby, 2006; Maddox et al., 2004; Rehder and Hoffman; Smith et al.). This is unexpected, given that the design of the categorisation phase of this study was comparable to previous studies, in terms of the number of stimuli used, the number of times each stimulus was presented, and the number of dimensions comprising each category rule.

4.2   Inference

The second hypothesis was that LS functions would result in significantly faster response times and higher average scores during the Inference phase. There were three statistically significant results from the Inference phase:

- for all participants, significantly higher average scores for NLS functions in the fourth trial
- for the Top Performers' group, significantly higher average scores for LS functions in the second trial
- for all participants, significantly faster response times for LS functions in the first trial.

These results do not provide strong support for the linear separability explanation for Johnson-Laird's findings (Johnson-Laird and Savary, 1996; Johnson-Laird and Savary, 1999). The fact that NLS Inference rates were significantly *higher* than LS in the fourth trial are counter to the hypothesis, and cannot easily be explained. However, there were a number of differences between this experiment, and the standard paradigm used by Johnson-Laird. For instance, this study was considerably longer and more repetitive; this may have left participants fatigued and unable to concentrate. Numerous participants expressed their frustration at the repetitive nature of the task. This disengagement also appears to be reflected in the declining Inference scores across the study. Even in the Top Performers group, accuracy scores for LS functions dropped from .96 on the first trial, to .59 on the fifth trial.

In addition, it may be that participants were unable to make the link between Categorisation and Inference. For instance, as discussed previously (see Footnote 9), one

participant showed a relatively normal learning curve during Categorisation, yet answered "don't know" to every Inference question. Finally, given that the Categorisation data suggested that some participants could not differentiate between LS and NLS categories while learning the rules, it is not unexpected that this lack of differentiation would extend to Inference.

4.3   Suggestions for future research

As there were a small number of significant differences between LS and NLS categorisation, the linear separability explanation should not be dismissed without further investigation. There are a number of directions this could take.

The first option for further study is a modified version of the current experiment. Results from the current study suggest that participants became fatigued and disengaged, due to the repetition. The clearest evidence of a linear separability effect occurred in the second trial, but performed decreased subsequently. If the experimental design was shortened to only one or two categorisation blocks (rather than the eight used in the current study) sequences, followed by an inference sequence, clearer separability effects may be evident. A variation on this may be to force participants to respond at a particular speed, rather than allowing them to respond at their own pace, to see if this produced any variations in response rates.

Another option for future study would be to use an experimental design more similar to some of the categorisation and inference studies discussed in Section 1.3 (Ashby et al., 2001; Ell and Ashby, 2006; Maddox et al., 2004; Rehder and Hoffman, 2005; Smith et al., 2011), including using overlapping categories, and testing on unique stimuli rather than novel. The primary aim of this proposed study would be to provide further evidence on the relationship between categorisation, inference, and separability. Examining the linear separability explanation for Johnson-Laird's work would be a secondary aim.

A third option for future study is a closer replication of Johnson-Laird's work. This would involve using the logical functions from this study, and converting them to word problems, similar to those used by Johnson-Laird. For example, one of the LS functions used in this study was B (C OR A). This could be changed into a word problem, like:

> At least one of the following statements is true, and possibly both.
> 1. You have the peanuts and the almonds
> 2. You have the peanuts and the walnuts
> Suppose you have the almonds. Is it possible for you to have the walnuts?

These problems could be generated either as Categorisation or Inference problems. This gives rise to a 2 x 2 experimental design, testing the effects of separability (LS vs. NLS) and type of problem (Categorisation vs. Inference). An additional factor might be to test the impact of problems that require falsification vs. problems that do not require falsification, as Johnson-Laird (Johnson-Laird and Savary, 1996; Johnson-Laird and Savary, 1999) suggests this affects the likelihood that people will solve problems correctly. There is some

overlap between falsification and linear separability, for instance, functions containing an XOR always require falsification, and are generally NLS, but not all NLS functions contain an XOR. It would be interesting to test if NLS problems that did not require falsification (if such problems exist) result in different average scores than NLS functions that did require falsification.

In this proposed study, each problem would be presented only once, consistent with Johnson-Laird's experimental designs. Results would provide further evidence on whether or not Johnson-Laird's findings are better explained by a separability effect than by the mental models theory. It would also help address the current lack of research on the difference between inference for LS and NLS functions (Markman and Ross, 2003; Yamauchi et al., 2002; Yamauchi and Markman, 1998).

A fourth option for further study is a replication of the current study with the addition of psychophysiological measures of workload. This is part of a growing program of work in DSTO. One area of interest to this research group is differentiating between low and high workload tasks, and tasks where the level of engagement varies. It is possible that categorisation and comprehending NLS and LS functions is an appropriate task for this work program.

Finally, it is important to identify the military implications of this work. In their study, Sparkes and Huf (2003) used versions of Johnson-Laird's problems, with military terms and concepts, e.g.:

> Only one of the following statements about a road convoy is true:
> (1) There is an Armoured Personnel Carrier in the convoy or there is a Tank in the convoy or both
> (2) There is a Mine Clearance Vehicle in the convoy and a Tank in the convoy
> Is it possible for there to be an Armoured Personal Carrier and a Tank in the convoy?

Anecdotal discussions with military personnel[10] suggest that military intelligence reports are not usually presented in such a format. However, there may be other areas of military decision making problems, where "either-or" problems do exist. These areas may be identified in future research. In addition, Sparkes and Huf (2003) suggested that by examining fundamental processes in decision-making, this area of research could provide two potential benefits to the military. Firstly, the results could be used to improve the speed of commanders' decision-making, allowing our forces to respond faster than an adversary. Secondly, the results could be used to improve the quality of decision support technology.

Sparkes and Huf's (2003) findings showed that military participants responded more quickly, but not more accurately, than civilian participants. This suggests that further improvements in the speed of decision making may not be as important as improvements

---

[10] Such as at the 2012 Pacific Armies Management Seminar, where this study was presented in poster format.

in the accuracy of decision making. However, the work may still have application in improving decision support technology. Another option is found in DSTO's ongoing program of work on complex adaptive decision-making. This work examines cognitive biases and reasoning errors, including identifying them, and providing training to mitigate against their effects (Grisogono and Radenovic, 2007). The Johnson-Laird paradigm is a classic example of a reasoning error, and could serve as an exemplar for identifying the conditions under which people are more or less likely to succumb to it.

In conclusion, the aim of this study was to examine if a linear separability effect was a plausible explanation for Johnson-Laird's findings. Only limited evidence in support of this explanation was found. However, it is possible this was due to methodological limitations, and so the linear separability explanation should not be discounted without further research. This could take the form of a closer replication of Johnson-Laird's experimental paradigm, draw on other cognitive experimentation paradigms, or integrate other research areas such as psychophysiology and complex adaptive decision-making. In order for this work to have military applications, it is important that steps are taken to identify real-world implications and analogues.

5. Acknowledgements

6. References

Ashby, F. G. and Maddox, W. T. (2005) Human category learning. *Annual Review of Psychology* **56** 149-178

Ashby, F. G., et al. (2001) Suboptimality in human categorization and identification. *Journal of Experimental Psychology: General* **130** 77-96

Barres, P. and Johnson-Laird, P. (2003) On imagining what is true (and what is false). *Thinking & Reasoning* **9** (1) 1-42

Barrouillet, P. and Lecas, J. F. (2000) Illusory inferences from a disjunction of conditionals: A new mental models account. *Cognition* **76** (2) 167-173

Blair, M. and Homa, D. (2001) Expanding the search for a linear separability constraint on category learning. *Memory & Cognition* **29** (8) Dec 1153-1164

Ell, S. W. and Ashby, F. G. (2006) The effects of category overlap on information-integration and rule-based category learning. *Attention, Perception, & Psychophysics* **68** (6) 1013-1026

Goodwin, G. P. and Johnson-Laird, P. (2010) Conceptual illusions. *Cognition* **114** (2) 253-265

Goodwin, G. P. and Johnson-Laird, P. N. (2011) Mental models of Boolean concepts. *Cognitive Psychology* **63** 34-59

Gregg, J. R. (1998) *Ones and zeros: Understanding Boolean algebra, digital circuits, and the logic of sets*. New York, IEEE Press

Grisogono, A. M. and Radenovic, V. (2007) The Adaptive Stance - steps towards teaching more effective complex decision-making. In: *Asia Pacific Complex Systems Conference,* Queensland, Australia: 2-5 July

Hopkins, W. G. *A scale of magnitude for effect size statistics*. (2002) [Accessed 2014 8 January]; Available from: http://www.sportsci.org/resource/stats/effectmag.html.

Johnson-Laird, P. N. (2010) Mental models and human reasoning. *Proceedings of the National Academy of Sciences* **107** (43) 18243-18250

Johnson-Laird, P. N., Byrne, R. M. J. and Girotto, V. (2009) The mental model theory of conditionals: A reply to Guy Politzer. *Topoi* **28** (1) 75-80

Johnson-Laird, P. N. and Savary, F. (1996) Illusory inferences about probabilities. *Acta Psychologica* **93** (1-3) 69-90

Johnson-Laird, P. N. and Savary, F. (1999) Illusory inferences: A novel class of erroneous deductions. *Cognition* **71** (3) 191-229

Khemlani, S. and Johnson-Laird, P. (2009) Disjunctive illusory inferences and how to eliminate them. *Memory & Cognition* **37** (5) 615-623

Maddox, W. T., Filoteo, J. V. and Hejl, K. D. (2004) Category number impacts rule-based but not information-integration category learning: Further evidence for dissociable category-learning systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **30** (1) 227

Markman, A. B. and Ross, B. H. (2003) Category use and category learning. *Psychological Bulletin* **129** (4) 592

Minsky, M. and Papert, S. (1972) *Perceptrons: An introduction fo computational geometry*. Cambridge, MA, The MIT Press

NHMRC (2007) *National Statement on Ethical Conduct in Human Research* Canberra, ACT

Rehder, B. and Hoffman, A. B. (2005) Eyetracking and selective attention in category learning. *Cognitive Psychology* **51** (1) 1-41

Santamaría, C. and Johnson-Laird, P. (2000) An antidote to illusory inferences. *Thinking & Reasoning* **6** (4) 313-333

Smith, J. D., Coutinho, M. V. C. and Couchman, J. J. (2011) The learning of exclusive-or categories by monkeys (*Macaca mulatta*) and humans (*Homo sapiens*). *Journal of Experimental Psychology: Animal Behavior Processes* **37** (1) 20

Sparkes, J. and Huf, S. (2003) *Mental models theory and military decision-making: A pilot experimental model [DSTO-GD-0368].* Department of Defence, Editor. Edinburgh, SA. Department of Defence,.

Temby, P., et al. (2005) Sharp shooting in fuzzy fields: Effects of image clarity on virtual environments. In: *SimTecT 2005,* Sydney, NSW

Whitney, S. (2013) *Literature review on mental models and linear separability (DSTO-GD-0741).* Edinburgh, SA. Defence Science and Technology Organisation.

Yamauchi, T., Love, B. C. and Markman, A. B. (2002) Learning nonlinearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **28** (3) May 585-593

Yamauchi, T. and Markman, A. B. (1998) Category learning by inference and classification. *Journal of Memory and Language* **39** (1) 124-148

Appendix A: Analysis of linear separability of Johnson-Laird's original problems

As discussed in Section 1.4.1, our analysis of Johnson-Laird's problems has identified a trend where problems that are difficult to are NLS, and problems that are easier to solve are LS. In this Appendix, we work through some examples from Johnson-Laird's previous research (Goodwin and Johnson-Laird, 2010; Goodwin and Johnson-Laird, 2011; Johnson-Laird and Savary, 1996; Johnson-Laird and Savary, 1999; Khemlani and Johnson-Laird, 2009; Santamaría and Johnson-Laird, 2000). We convert each problem to a Boolean algebra statement, generate a truth table, and graph the solutions. We then identify whether the problem is LS or NLS, and report the number of participants answering the problem correctly in the original study.

A.1.   Overview of logical principles

In order to explain this analysis in detail it is essential to cover some of the basics of logical reasoning. To begin, consider the following logical problem.

> If the server is full, then memory is busy.
> The server is full. What, if anything, can be deduced about memory?

It is reasonably straightforward to deduce that the answer to this problem is that if the server is full, then it follows that memory must be busy. However, what if the server is not full? What can be deduced about the state of memory in this situation?

The problem above is an example of a logical statement of the form *if A then B*. Under formal logical rules[11], this means that B, known as the *consequent*, always follows in the presence of A, known as the *antecedent*. The relationship between A and B is of a type known as a *conditional*, where the presence of A implies the presence of B.

One way of representing if A, then B is through a truth table. Table 1 shows the possible values (true or false) for A and B, and the resultant values for the statement if A, then B. The first two lines are reasonably intuitive; if A and B are both true, then the statement is true, and if B is false, then the statement is false. The third and fourth lines demonstrate an important logical principle: if the antecedent is false, then the consequent can be true or false, and the statement will still be logically true. Using the example above, if the server is *not* full, then memory can be either busy or not busy without the statement being logically false.

---

[11] *Modus ponens;* see, for instance http://en.wikipedia.org/wiki/Modus_ponens for more detail.

Table 1: Truth table for If A, then B

A	B	If A, then B
True	True	True
True	False	False
False	False	True
False	True	True

The logical relationships relevant to understanding the analysis of Johnson-Laird's problems are:

- *AND* – all values are true, e.g. *A AND B* means that both A and B are true.

- *OR* – one, some, or all values are true, e.g. *A OR B* is true when A is true, B is true, and when A and B is true.

- *Exclusive OR* (*XOR*) – only one value is true, e.g. *A XOR B* means that either A or B, but not both, is true.

- *NOT*, ⌐ – this means that a value is not true.

For more detail on these operators, or the principles of Boolean algebra underpinning them, the reader is referred to Gregg (1998) or Whitney (2013).

A.2.   Johnson-Laird and Savary (1996), Experiment 1

In their first study, Johnson-Laird and Savary (1996) used four problems[12]. These were:

Problem 1:
Only one statement about a hand of cards is true:
(1) There is a King or Ace or both.
(2) There is a Queen or Ace or both.
Which is more likely, King or Ace?

Problem 2:
Only one statement about a hand of cards is true:
(1) If there is a King in the hand, there is an Ace in the hand.
(2) If there is a Queen in the hand, then there is an ace in the hand.
Which is more likely, King or Ace?

Problem 3:
If there is a King in the hand, then there is an Ace in the hand. Which is more likely, King or Ace?

---

[12] Note that Johnson-Laird and Savary (1996) used different cards for each of their problems rather than repeating Ace, King, and Queen. However, to avoid confusion we have kept the terms consistent for this analysis.

Problem 4:
If there is a King or a Queen in the hand, then there is an Ace in the hand.
Which is more likely, King or Ace?

The first two problems required falsification of mental models, (using Johnson-Laird's terminology), while the second two did not. The results, summarised in Table 2, are consistent with the mental models theory. That is, the problems requiring falsification were answered correctly by a significantly lower percentage of participants than the problems that did not require falsification.

Table 2: Summary of correct answers and percentage of participants answering correctly

	Correct answer	% who answered correctly
Problem 1	King	21%
Problem 2	King	13%
Problem 3	Ace	62%
Problem 4	Ace	79%

In the following sections, we write each of these problems as Boolean equations, create truth tables, and graph solutions in order to identify if the problems are LS or NLS.

A.2.1 Problem 1

Only one statement about a hand of cards is true:
(1) There is a King or Ace or both.
(2) There is a Queen or Ace or both.
Which is more likely, King or Ace?

This problem, which is also discussed in detail in Section 1.4, is represented by the equation (King OR Ace) XOR (Queen OR Ace).

As the Ace occurs on both sides of the XOR, it is removed from the equation, which simplifies to King XOR Queen. This produces the truth table shown in Table 3, and the graphical solution shown in Figure 24.

Table 3: Truth table for the equation King XOR Queen

King	Queen	King XOR Queen
False	False	False
False	True	True
True	False	True
True	True	False

As discussed in the body of the text, the fact that the Ace must be removed from the equation means that it can never occur, and hence that the King is more likely to occur. Figure 24 also demonstrates that this is a NLS problem, due to the inability to drawn a single straight line separating true from false answers.

	True	False
King		
⌐King	False	True
	⌐Queen	Queen

Figure 24: Solutions for the equation King XOR Queen

A.2.2 Problem 2

Only one statement about a hand of cards is true:
(1) If there is a King in the hand, there is an Ace in the hand.
(2) If there is a Queen in the hand, then there is an ace in the hand.
Which is more likely, King or Ace?

This problem is represented by the equation (King AND Ace) XOR (Queen AND Ace). The Ace must be removed from both sides of the equation, for the same logical reasons as in Problem 1. This produces the same truth table and graphical solution as the preceding problem. Again, this is a NLS problem, where the King is more likely to occur.

A.2.3 Problem 3

If there is a King in the hand, then there is an Ace in the hand. Which is more likely, King or Ace?

This problem is represented by the equation (King AND Ace) OR Ace. The "OR Ace" term is included because as discussed above, if the antecedent is false, then the consequent can be true or false, and the statement will still be logically true. That is, if King is not present, the Ace can still be present.

The equation gives the following truth table (Table 2), and the graphical solution shown in Figure 25. It is clear that there are only two logically possible answers, Ace and (King AND Ace). In these answers, the Ace occurs twice and the King only once, therefore the correct answer is Ace. It is also clear from looking at the figure that a line can be drawn differentiating between true and false answers (indicated by the dotted line), hence this is a LS problem.

Table 4: Truth table for King OR Ace

King	Ace	King AND Ace	(King AND Ace) OR Ace
False	False	False	False
False	True	False	True
True	False	False	False
True	True	True	True

Figure 25: Solutions for the problem (King AND Ace) OR Ace.

A.2.4 Problem 4

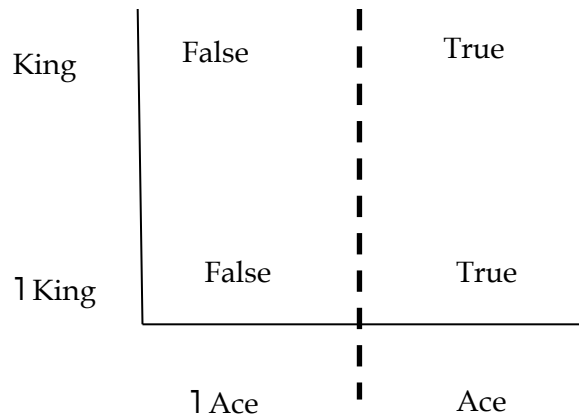*If there is a King or a Queen in the hand, then there is an Ace in the hand. Which is more likely, King or Ace?*

This problem is represented by the equation (King OR Queen) AND Ace. It produces the truth table shown in Table 5. The table shows that for the three correct answers (Queen AND Ace; King AND Ace; Queen, King, AND Ace), the Ace occurs in all three answers while the King occurs in only two answers. Therefore the correct answer is that the Ace is more likely to occur.

Table 5: Truth table for (King OR Queen) AND Ace

King	Queen	Ace	(King OR Queen)	(King OR Queen) AND Ace
False	False	False	False	False
False	False	True	False	False
False	True	False	True	False
False	True	True	True	True
True	False	False	True	False
True	False	True	True	True
True	True	False	True	False
True	True	True	True	True

This problem contains three terms rather than two. To demonstrate that it is linearly separable requires drawing a graph with three axes, and drawing a two-dimensional plane to show the separation between true and false answers. This is shown in Figure 26. In the figure, starbursts at intersections of the three axes represent the three possible true answers (King AND Ace; Queen AND Ace; King AND Queen AND Ace). Note that the z-axis (Queen) reads from back to front. This is unconventional, but this orientation of axes most clearly shows the separation.
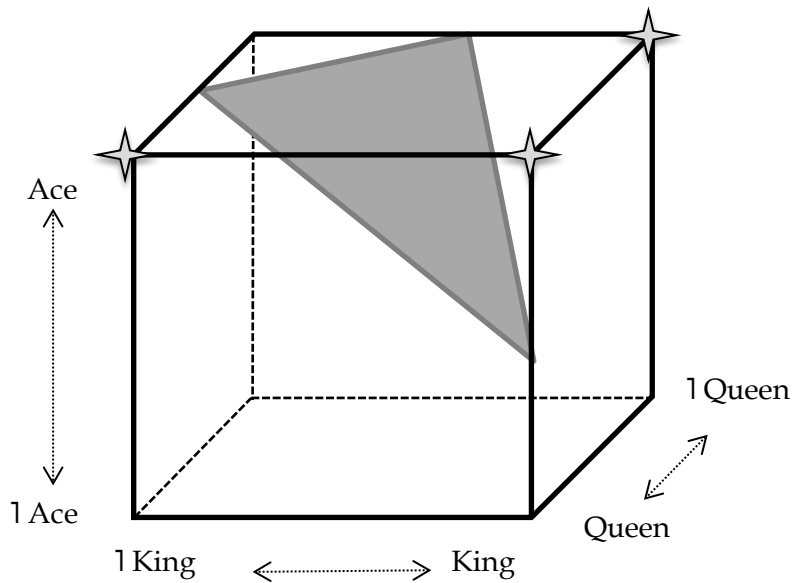

Figure 26: Solution for (King OR Queen) AND Ace

The analysis of these four problems clearly demonstrates that the two LS problems were solved correctly by the majority of participants, while the two NLS problems were solved correctly by only a small percentage of the participants. One potential confound is that the first three problems only used two terms, while the fourth problem contained three terms. We addressed this in our study by ensuring that all problems contained three terms.

A.3.   Analysis of other problems used by Johnson-Laird

In this section, we analyse some problems used by Johnson-Laird in subsequent studies. This is further demonstration that problems participants struggle to answer tend to be NLS, while the problems they answer easily tend to be LS. We have chosen two problems from each study, one that received a high percentage of correct responses, and one that received a low percentage of responses. The next two problems are from Johnson-Laird and Savary (1999).

A.3.1 Problem 5

There is a king in the hand and there is not an ace in the hand, or else there is
an ace in the hand and there is not a king in the hand.
There is a king in the hand.
What, if anything, follows?

This problem is represented by the equation (King AND NOT-Ace) XOR (Ace AND NOT-King), and produces the following truth table, and the graphical solution shown in Figure 27. It is clear from the figure that this problem is NLS as the true and false answers cannot be separated by a straight line. All participants failed to solve this problem correctly.

Table 6: Truth table for (King AND NOT-Ace) XOR (Ace AND NOT-King)

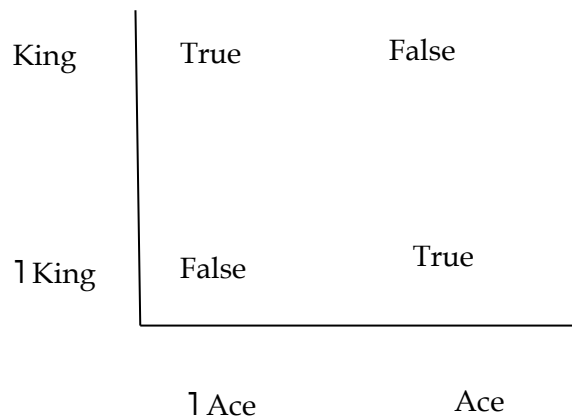| King	Ace	King AND NOT-Ace	Ace AND NOT-King	(King AND NOT-Ace) XOR (Ace AND NOT-King)
False	False	False	False	False
False	True	False	True	True
True	False	True	False	True
True	True	False	False	False

Figure 27: Solutions for (King AND NOT-Ace) XOR (Ace AND NOT-King)

A.3.2 Problem 6

If there is a king in the hand then there is an ace in the hand, or else there is not
a king in the hand.
There is a king in the hand.
What, if anything, follows?

This problem is represented by the equation (King AND Ace) XOR (NOT-King), which produces the truth table shown in Table 7, and the graphical solution shown in YYY. The figure clearly shows that this problem is LS. This problem was solved correctly by 100% of participants in Johnson-Laird and Savary (1999).

Table 7: Truth table for (King AND Ace) XOR (NOT-King)

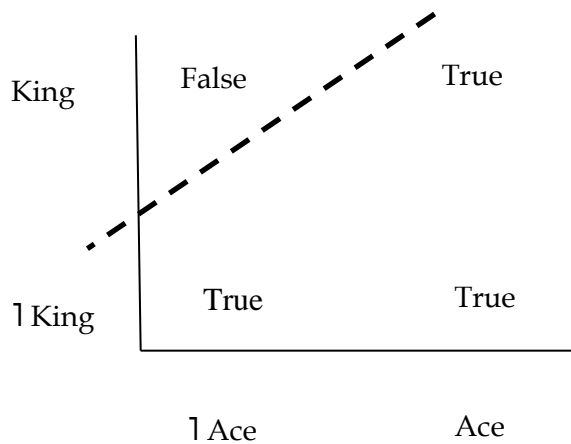| King	Ace	King AND Ace	NOT-King	(King AND Ace) XOR (NOT-King)
False	False	False	True	True
False	True	False	True	True
True	False	False	False	False
True	True	True	False	True

Figure 28: Solutions for (King AND Ace) XOR (NOT-King)

A.3.3 Problem 7

This problem and the next were used in Santamaría and Johnson-Laird (2000).

> Only one of the two following assertions is true about John:
> (1) John is a lawyer or an economist, or both.
> (2) John is a sociologist or an economist, or both.
> He is not both a lawyer and a sociologist.
> Is John an economist?

This problem is represented by the equation (Lawyer OR Economist) XOR (Sociologist or Economist). It is clear that this equation is of the form (A OR B) XOR (B OR C), which is the same form used in Problem 1. Hence, the truth table and figure for this problem is the same as for Problem 1, as is the conclusion that this is an NLS problem. This problem was solved correctly by 6% of participants (Santamaría and Johnson-Laird, 2000).

A.3.4 Problem 8

Only one of the two following assertions is true about John:
(1) John is a lawyer or an economist, or both.
(2) John is a sociologist or an economist, or both.
He is not a lawyer and he is not an economist.
Is John a sociologist?

The equation for this problem is identical to the equation used for Problem 1 and Problem 7. It takes the form (A OR B) XOR (B OR C). The term that occurs on both sides of the XOR – in this case B, or economist – is removed from the equation, leaving A XOR C, or Lawyer XOR Sociologist. This produces the truth table shown in Table 8, and the graphical solutions shown in Figure 29. Although this problem is NLS, it was solved correctly by 100% of participants in Santamaría and Johnson-Laird (2000). This is an exception to our observation that NLS problems are more difficult to solve.

Table 8: Truth table for Lawyer XOR Economist

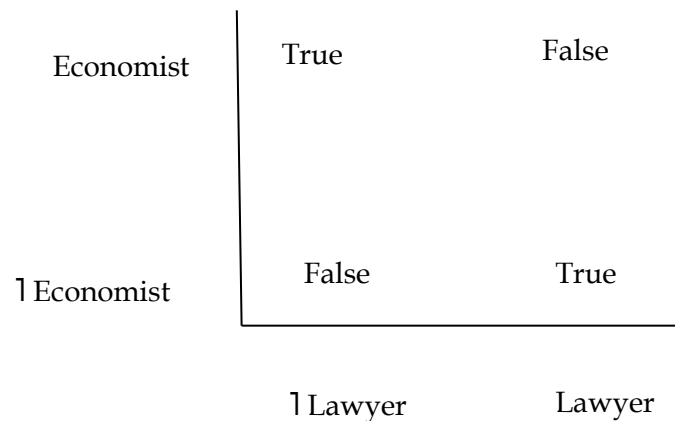| Lawyer	Economist	Lawyer XOR Economist
False	False	False
False	True	True
True	False	True
True	True	False

Figure 29: Solutions for Lawyer XOR Economist

A.3.5 Problem 9

This problem, and the next were used by Khemlani and Johnson-Laird (2009).

Suppose that only one of the following assertions is true:
(1) You have the mints.
(2) You have the gumballs or the lollipops, but not both.
Also, suppose you have the mints. What, if anything, follows? Is it possible that you also have either the gumballs or the lollipops? Could you have both?

This problem can be expressed as Mints XOR (Gumballs OR Lollipops). This produces the following truth table, and the graphical solution shown in Figure 30. The figure indicates that it is impossible to draw a two-dimensional plane separating the true from false answers. This NLS problem was solved correctly by only 17% of participants.

Table 9: Truth table for Mints XOR (Gumballs or Lollipops)

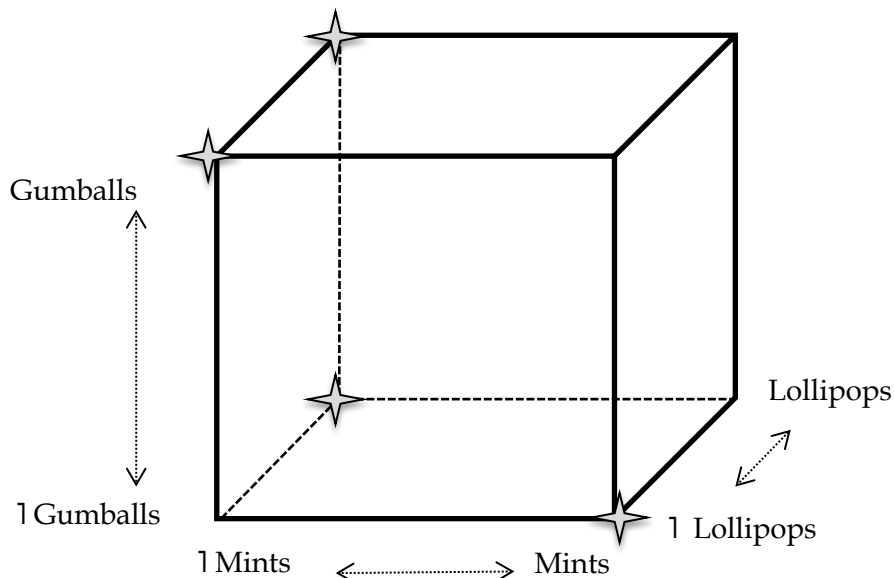| Mints	Gumballs	Lollipops	(Gumballs OR Lollipops)	Mints XOR (Gumballs OR Lollipops)
False	False	False	False	False
False	False	True	True	True
False	True	False	True	True
False	True	True	True	True
True	False	False	False	True
True	False	True	True	False
True	True	False	True	False
True	True	True	True	False

Figure 30: Solutions for Mints XOR (Gumballs OR Lollipops)

A.3.6 Problem 10

Suppose that at least one of the following assertions is true, and possibly both:
(1) You have the marshmallows.
(2) You have the truffles or the Jolly Ranchers, and possibly both.
Also, suppose you have the marshmallows. What, if anything, follows? Is it possible that you also have either the truffles or Jolly Ranchers? Could you have both?

This can be expressed in the equation Marshmallows OR Truffles OR Jolly Ranchers. This produces the truth table shown in Table 10, and the solutions shown in Figure 31. The figure indicates that it is possible to draw a two-dimensional plane that separates true from false answers, hence this is a LS problem. This problem was answered correctly by 100% of participants. This problem is also discussed in the body of the report in Section 1.4.1 and Footnote 5.

Table 10: Truth table for Marshmallows OR Truffles OR Jolly Ranchers

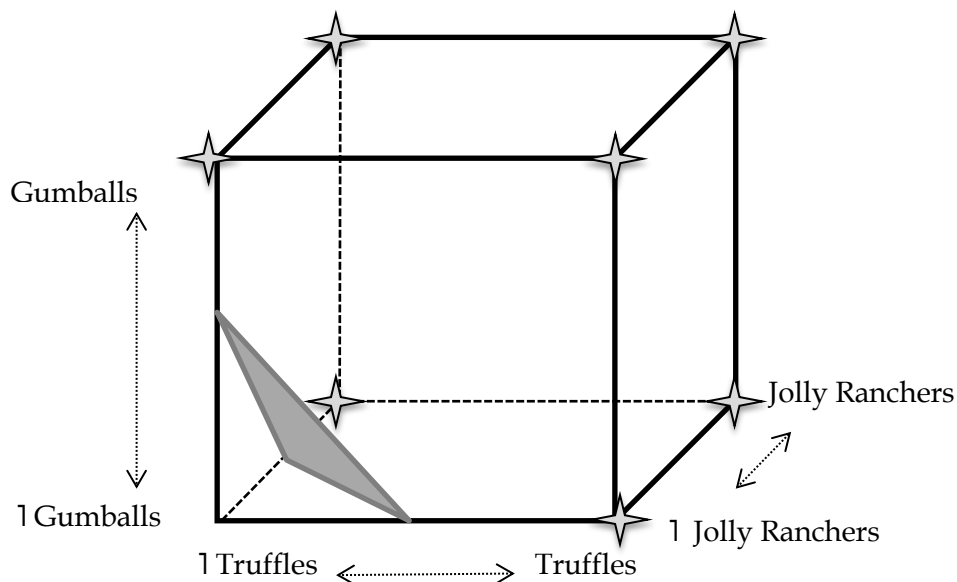| Marshmallows	Truffles	Jolly Ranchers	Marshmallows OR Truffles OR Jolly Ranchers
False	False	False	False
False	False	True	True
False	True	False	True
False	True	True	True
True	False	False	True
True	False	True	True
True	True	False	True
True	True	True	True

Figure 31: Solutions for Gumballs OR Truffles OR Jolly Ranchers

A.3.7 Problem 11

This problem, and the next problem, are taken from Goodwin and Johnson-Laird (2010)[13]

(A AND B), XOR (NOT-A AND B)

This produces the truth table shown in Table 11, and the solutions shown in Figure 32. This problems is LS, and was solved correctly by 95% of participants.

Table 11: Truth table for (A AND B), XOR (NOT-A AND B)

A	B	(A AND B)	(NOT-A AND B)	(A AND B) XOR (NOT-A AND B)
False	False	False	False	False
False	True	False	True	True
True	False	False	False	False
True	True	True	False	True
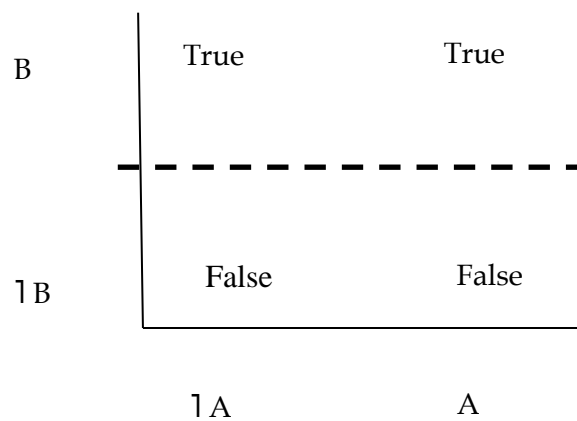
Figure 32: Solutions for (A AND B) XOR (NOT-A AND B)

A.3.8 Problem 12

(A AND B) XOR (NOT-A AND NOT-B) – 17%

This produces the following truth table, and the solutions shown in Figure 33. This problem is NLS, and was solved correctly by only 17% of participants.

---

[13] In their study, Goodwin and Johnson-Laird use "or else" instead of XOR. However, the meaning is the same.

Table 12: Truth table for (A AND B) XOR (NOT-A AND NOT-B)

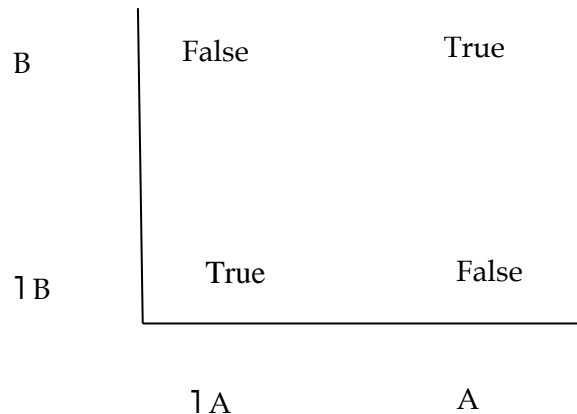| A	B	(A AND B)	(NOT-A AND NOT- B)	(A AND B) XOR (NOT-A AND NOT-B)
False	False	False	True	True
False	True	False	False	False
True	False	False	False	False
True	True	True	False	True

Figure 33: Solutions for(A AND B) XOR (NOT-A AND NOT-B)

A.3.9 Problem 13

This problem and the following problem are from Goodwin and Johnson-Laird (2011).

(A AND NOT-B) OR (B AND C)

This produces the truth table below, and the solutions shown in Figure 34. It is clear from the figure that this problem is NLS as it is impossible to draw a two-dimensional plane separating true from false answers. This problem was solved correctly by only 48% of participants.

Table 13: (A AND NOT-B) OR (B AND C)

A	B	C	(A AND NOT-B)	(B AND C)	(A AND NOT-B) OR (B AND C)
False	False	False	False	False	False
False	False	True	False	False	False
False	True	False	False	False	False
False	True	True	False	True	True
True	False	False	True	False	True
True	False	True	True	False	True
True	True	False	False	False	False
True	True	True	False	True	True

Figure 34: Solutions for (A AND NOT-B) OR (B AND C)

A.3.10 Problem 14

A AND NOT-B

This produces the truth table shown in Table 14, and the solutions shown in Figure 35. It is clear that this is a LS problem, and it was solved correctly by 100% of participants

Table 14: Truth table for A AND NOT-B

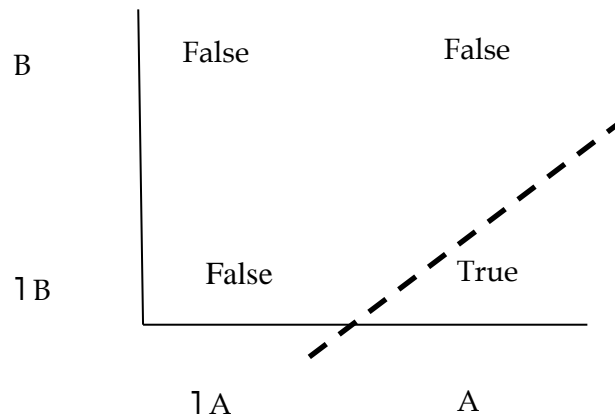| A	B	(A AND NOT-B)
False	False	False
False	True	False
True	False	True
True	True	False

Figure 35: Solutions for A AND NOT-B

A.4.   Conclusion

We have analysed 14 problems from 6 studies conducted by Johnson-Laird and colleagues. As summarised in Table 15, LS problems tended to produce high rates of correct responses, frequently at or close to ceiling. In contrast, the majority of NLS problems produced poor rates of correct responses. With the exception of Problem 8, the NLS problems were answered correctly by fewer than half the participant.

Table 15: Summary of problem separability and percentage of participants correctly answering

LS problems	% correct	NLS problems	% correct
Problem 3	62	Problem 1	21
Problem 4	79	Problem 2	13
Problem 6	100	Problem 5	0
Problem 10	100	Problem 7	7
Problem 11	95	Problem 8	100
Problem 14	100	Problem 9	17
		Problem 12	17
		Problem 13	48

We have not analysed all problems used by Johnson-Laird. We believe that, even if the results were consistent with the analysis we have conducted so far, the impact of other potentially confounding factors cannot be ruled out. For instance, some of Johnson-Laird's problems contain two terms, others contain three terms. In some of the problems containing three terms it is necessary to consider all three terms, whereas in other problems a term can be removed from the equation. This may affect the difficulty of solving the problem.

In addition, across studies there is variance in the level of clarity provided to participants about the exclusionary nature of the XOR function and "or else" term, or variance in the clarity of the problem statement. As discussed in Footnote 6, p8 some participants may

have trouble understanding which terms an XOR or "or else" refers to. In some of Johnson-Laird's studies, this is made explicit, while in others it is not emphasised. This introduces another potentially confounding factor, the level of concreteness of the problems. Problems where the function of the XOR or "'or else" is clearer tend to be more concrete (such as Problems 7-10), while problems where the function is less clear tend to be more abstract (such as Problems 11 and 12). Again, this may affect the difficulty of solving the problem.

Based on the analysis we have conducted, we believe that the tendency for separability to affect that solvability of problems is strong enough to warrant the specific hypotheses tested in this study. This study was designed to overcome the potentially confounding factors. As discussed in more detail in the body of the report, each problem contained three terms (which all needed to be considered to solve it correctly), and all problems were presented in an identical format with identical levels of clarity.

Appendix B: List of functions used in the study

B.1. Linearly separable functions

Practise function:
1. C AND (B OR A)

Test functions:
1. B (C OR A)
2. B (¬A¬C OR A)
3. B (A¬C OR C)
4. ¬B(¬A OR C)
5. ¬B(¬A OR ¬C)
6. ¬A(C OR B)

B.2. Nonlinearly separable functions

Test functions:
1. (¬ABC) OR (A¬BC) OR (AB¬C)
2. (BC) OR (A¬B¬C)
3. C(A XOR B) OR (AB¬C)
4. (B¬C) OR (A¬BC)
5. (¬A¬BC) OR A(B XOR C)
6. (¬B¬C) OR (¬ABC)

Appendix C: Onscreen instructions

Information

- You will be presented with a series of shape combinations consisting of a square, circle and triangle.
  Each shape can be either shaded or unshaded.

- A hypothetical light exists, so that some combinations of the shaded and/or unshaded shapes will turn the light on.
  Your main task will be to learn which combinations of shapes and shading activate the light.

- Your task is to examine each combination and decide whether it will turn the light 'On' or remain 'Off'.

- Use the mouse to indicate whether you believe the combination will turn the light On or Off.
  Whether you are correct or incorrect will be revealed immediately following your response.

- You will have ten seconds to view the combination, then it will disappear, but responses can still be entered after this time.

- This will be followed by a set of questions without feedback.
  You will be required to give information on a missing shape from a combination using multiple choice format.

- For some combinations there is only one answer to the colour of the missing shape ('shaded' or 'unshaded').
  But for others the shape can be either shaded or unshaded and still produce the same outcome in the light (answer 'either').

- You will now be presented with an example set of shapes following the format of the remainder of the experiment.
  This will be followed by the set of questions to answer.

[Continue]

Appendix D: Post experimental survey

Thank you for participating in the experiment. Please answer the following questions.

1. You solved two sets of light switch problems. Did you find (please circle one):
a) The first set of problems was easier to solve
b) The second set of problem was easier to solve
c) Both sets of problems were equally easy to solve

2. Which response best describes how you solved the problems? Did you:
a) Try to memorise the correct answers
b) Try to deduce the rule behind the problems
c) Guess which answers were correct
d) Other – please describe
_____
_____
_____

3. How confident are you that you solved the problems correctly?
a) Not at all confident
b) Moderately confident
c) Very confident

Do you have any other comments about the experiment?

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION DOCUMENT CONTROL DATA		1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)
2. TITLE Linear Separability in Categorisation and Inference: A Test of the Johnson-Laird Falsity Model	3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION) Document (U) Title (U) Abstract (U)	
4. AUTHOR(S) Susannah J. Whitney, George Galanis and Armando Vozzo	5. CORPORATE AUTHOR DSTO Defence Science and Technology Organisation 506 Lorimer St Fishermans Bend Victoria 3207 Australia	

6a. DSTO NUMBER DSTO-TR-2935	6b. AR NUMBER AR-015-846	6c. TYPE OF REPORT Technical Report	7. DOCUMENT DATE January 2014	
8. FILE NUMBER 2012/1236963/1	9. TASK NUMBER LOD.3 - LHS ER&D	10. TASK SPONSOR CLD	11. NO. OF PAGES 51	12. NO. OF REFERENCES 34

DSTO Publications Repository http://dspace.dsto.defence.gov.au/dspace/	14. RELEASE AUTHORITY Chief, Land Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT *Approved for public release*   OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111
16. DELIBERATE ANNOUNCEMENT No Limitations

17. CITATION IN OTHER DOCUMENTS	Yes

18. DSTO RESEARCH LIBRARY THESAURUS Decision making, cognitive processes, experimental analysis

19. ABSTRACT
Johnson-Laird suggests that difficulties in problem solving can be explained by the mental models theory. This study tests linear seperability effects in categorisation and inference as an alternate explanation, hypothesising that categorisation and inference would be easier for linearly separable (LS) functions than nonlinearly separable (NLS). Thirty two participants were tested on one LS and one NLS functions over repeated trials. Results indicated that categorisation and inference were significantly more difficult for NLS functions, but only for the highest performing participants on some trials. Among poorer performing participants there were no significant differences between response rates and response times. The most likely explanations for these findings are the complexity and duration of the experiment, rather than lack of support for the linear separability hypothesis. Implications for the military and research communities and suggestions for future research are discussed.