

**DIVISION OF COMPUTER SCIENCE**

**Transfer of Non-logical Tendencies to Formal Reasoning**

**R J Vinter**

**Technical Report No. 252**

**July 1996**

# Transfer of Non-logical Tendencies to Formal Reasoning

R J Vinter

Division of Computer Science  
University of Hertfordshire

Tel: +44-01707-285122  
Email: R.Vinter@herts.ac.uk

July 1996

## Abstract

Previous psychological studies have shown that people are prone to systematic errors and biases when reasoning about natural language statements connected by logical operators such as: *if, and, or, not, all* and *every*. In the software engineering community, the use of formally defined notations based upon mathematical systems of logic are starting to gain favour over natural language based methods and it is generally believed that these will lead to improved reasoning. Although many formal notational constructs appear to share logical similarities with their natural language counterparts, they do not appear to be psychologically equivalent. This report describes a series of investigations aimed at determining whether people are inclined to use the same cognitive processes to reason about logical arguments irrespective of their linguistic context, and whether certain non-logical tendencies are ubiquitous across linguistic domains. In view of the fact that formal notations are commonly used for the development of safety-critical systems, this possible transfer of erroneous reasoning processes constitutes a genuine area for concern because developers who are unable to interpret or reason clearly about system specifications are likely to make the types of erroneous development decisions which have previously led to the production of defective systems.

## 1.0 Introduction

In the past, poorly written software specifications have led to developers making false assumptions and incorrect decisions which have had repercussions throughout the latter stages of software projects, causing the appearance of faults or anomalies in the system design or code produced. Within the software engineering community, it is widely believed that formal logic based notations could hold the key to overcoming some of these problems (Potter et al., 1991). This is perhaps reflected by the increasing acceptance of formal methods in industry and the gradual decline of traditional natural language based notations. Accordingly, computing science research appears to be concentrating its efforts on developing sleeker formal notations based on more complex grammars, perhaps with the short-term goals of simplifying program refinement and verification for the general practitioner. However, such research does not appear to be considering the social or psychological implications of using such notations. In particular, it has failed to examine the effects that the use of such notations has on their users' cognitive reasoning processes and has neglected the quite real possibility that the same non-logical tendencies which people exhibit when reasoning about implicit logic in natural language also transfer over into the domain of reasoning about explicit logic in formal specifications.

"It (logic) is justified in abstracting - indeed, it is under obligation to do so - from all objects of knowledge and their differences, leaving the understanding nothing to deal with save itself and its form."

Kant (reprinted in Smith, 1993, p. 18).

Within the software engineering community, it is generally believed that it must be easier to reason about arguments expressed in formal logic than natural language because logic abstracts away all extraneous, potentially distracting details and allows the reasoner to concentrate purely upon an argument's underlying form. However, any such theory might be refuted on at least two counts. Firstly, it appears to over-estimate logical deduction as a means of reasoning in comparison with other forms, such as those based on: induction, calculation, statistical probability or just plain intuition. In situations where only logical, deductive inferences can lead to correct conclusions, writers are powerless to prevent their readers from adopting illogical, non-deductive forms of reasoning. It would be reasonable to assume that, since most of the inferences that people make in everyday life are non-deductive, it is these methods to which they are most accustomed and most likely to adopt in situations where doubt arises. Secondly, the theory might be refuted on the grounds that it discounts the real possibility that such "extraneous details" may not in fact be distracting at all, but could actually help to promote sound reasoning. For evidence of this, one only has to witness the high rate of reasoners who failed to solve the abstract version of Wason's selection task (Wason, 1966) and the facilitatory effects brought about by realistic material in the task's different, but logically equivalent, guises (Griggs and Cox, 1982; Griggs and Jackson, 1990; Johnson-Laird and Wason, 1970; Wason and Shapiro, 1971). Perhaps above all, the results from such studies emphasise that the way in which a problem is presented can dramatically affect the ease with which people are able to solve it. Turner (1986) states that any writer who is concerned about the ways in which his work is articulated cannot lay down an intonation for his reader using natural language except where this is prescribed by the grammatical structure of what he writes and its context. Similarly, it is theorised that a software designer cannot lay down a system of reasoning for his readers except where this is implied by what he writes and its context. If the way in which a specification is written can facilitate or impair reasoning then, in order to minimise the potential for incorrect development decisions, it would appear essential that it is written in the clearest and most intuitive manner possible.

"Formal methods are no guarantee of correctness; they are applied by humans, who are obviously prone to error... System development is a human activity, and always will be. Software engineering will be prone to human whim, indecision, the ambiguity of natural language, and simple carelessness."

Bowen and Hinchey (1994a).

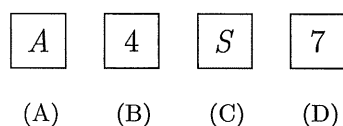
It is thought that much of the software engineering community's eagerness to switch to the formal approach might be founded on a major misconception: simply because a system's description is expressed in a precisely defined language does not mean that readers will necessarily adhere to the formal semantics of the language when interpreting it, nor adhere to the logic of its formal inference rules when reasoning about it. Previous computing science research does not appear to have considered the possibility that people may be susceptible to the same forms of reasoning errors in formal logic as those shown to be incited by natural language. Indeed, it is hypothesised that people use the same or similar higher-level, cognitive methods to reason about logical problems, irrespective of their linguistic form and that certain non-logical tendencies are ubiquitous across linguistic domains. Indeed, if formal notations can be shown to incite the same forms of reasoning errors exhibited by natural language then this should constitute a genuine reason for concern, especially in view of the fact that formal methods are regularly used in the development of safety-critical systems.

Over the past three decades, psychology has shown people to exhibit systematic biases and errors when reasoning about natural language statements containing words such as: *if, and, or, not, all* and *some* (Braine and O'Brien, 1991; Erickson, 1978; Johnson-Laird, 1977; Johnson-Laird and Tridgell, 1972; Lakoff, 1971; Newsstead et al., 1984). Despite obvious differences in their symbolic appearance, most formal logic based notations contain propositional connectives and quantifiers with roughly equivalent semantic definitions to those grammatical constructs identified as inciting non-logical reasoning tendencies in natural language:  $\Rightarrow$ ,  $\wedge$ ,  $\vee$ ,  $\neg$ ,  $\forall$  and  $\exists$ . It seems reasonable that an investigation aimed at discovering those properties of formal specifications that are particularly prone to inciting erroneous reasoning processes should begin with a study of similar grammatical constructs that have been proven to incite systematic reasoning errors in alternative linguistic domains.

## 2.0 Completed Experiments

### 2.1 Experiment 1: Reading, Writing, Understanding, Translating and Reasoning About Formal Specifications

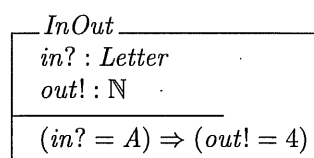
Design of the first experiment was oriented towards five cognitive activities deemed central to the formal specification process: reading, writing, understanding, translating and reasoning. Its main aim was to help refine the scope of the project and its original aims. Twelve computer scientists from the University of Hertfordshire with at least a basic understanding of the Z notation participated in the experiment: six staff and six students. Their mean age was 30 years and their mean level of experience with the Z notation was 31 months. The study had a repeated measures design, with all participants completing all tasks. Task sheets were taken away by participants and completed anonymously then returned to the experimenter via internal mail. Participants were asked to provide brief biographical details including: age, occupation, course, and details of their previous Z experience.



If there is an *A* on one side of the card then there is a 4 on the other.

Which cards would you need to turn over in order to determine whether the rule is true or false?

Figure 1: Wason's Abstract Selection Task.



- (A)  $in? = A$       (B)  $out! = 4$   
 (C)  $in? = S$       (D)  $out! = 7$

Which inputs and outputs would enable you to test whether 'InOut' is working correctly?

Figure 2: The Formalised Selection Task.

Wason's abstract selection task (Figure 1) is one of hypothesis testing and deductive reasoning based on conditional logic (Evans et al., 1993, p. 99-135; Wason, 1966). The first experimental task presented to participants was a variation on this task set within the context of a Z specification (Figure 2). The choices of responses given to participants corresponded to the  $p$ ,  $q$ ,  $\neg p$  and  $\neg q$  cases for a conditional rule of the form *if p then q*. In Wason's natural language based version of the task, the conditional rule *if p then q* is implicit, whereas in the formalised version, the conditional was presented in the form of an explicit logical implication statement. Participants needed to employ both the affirming *modus ponens* (MP) and denying *modus tollens* (MT) inferences in order to deduce the correct combination of responses: namely, the cases corresponding to  $p$  and  $\neg q$ .

Intuitively, it seems reasonable that participants with backgrounds in formal logic would be more likely to recognise the type of inference required to deduce the correct combination of responses to the formalised selection task because their recognition of the explicit conditional operator would ensure that they endorse only what follows logically. However, the observed results suggest that although every participant evaluated the  $p$  case as being relevant, none saw the relevance of the  $\neg q$  case. Hence, the 0% success rate for the formalised version of Wason's selection task was even lower than the 4% observed during Wason's early trials (Wason and Johnson-Laird, 1972, p. 182). Nevertheless, there are clear similarities between the patterns of results obtained from the two studies. Specifically, the frequency at which participants chose the  $p, q$  combination was highest, with selection of the  $p$  case alone being second highest in both studies. The high rate of  $p, q$  selections might be explained by the possibility that participants in both studies had succumbed to a strong form of "matching bias" whereby their responses were based purely on those terms explicitly mentioned in the conditional rule (Evans, 1972b; 1983). These participants' selections could therefore have been based on associative guesswork rather than logical deduction. The high rate of  $p$  selections and low rate of  $\neg q$  selections appears to shed some light on the relative difficulty with which people are able to draw MP and MT inferences in general. The clear correlations between the two studies suggest that Wason's findings do indeed carry over into the domain of formal specification and that, contrary to intuition, people do not necessarily find it easier to reason logically about conditionals in formal logic.

Given a text written in a foreign or technical language, it is generally believed that a reader will implicitly translate it into an appropriate form in their native language before attempting to reason about its contents. In this light, it seems important that people must be able to translate between formal expressions and equivalent natural language forms clearly and intuitively so that minimum distraction is caused when they are either interpreting or reasoning about formal specifications. The aim of the second experimental task was to test the intuitiveness of this translation process: Task 2a tested conceptual mappings from Z to English, whilst Task 2b tested them in the opposite direction.

Potter et al. (1991, p. 124) present the formal specification of a computerised library system containing four predicates. However, for the purposes of Task 2a, its fourth predicate was modified in order to oppose people's general conceptions of library systems (Figure 3). Participants were prompted to translate the specification's predicates into appropriate forms in natural English. Although it was predicted that most would succeed in offering English translations whose meanings coincided with those of the first three predicates, it was expected that participants' efforts in translating the complex fourth predicate would throw some light on the intuitiveness of conceptual mappings between formal logic and natural language.

Original fourth predicate:  $\forall r : readers \bullet \#(issued \triangleright \{r\}) \leq maxloans$   
 Logical meaning: "The number of books that any reader borrows must be less than or equal to the maximum number of loans allowed."  
 Revised fourth predicate:  $\neg \exists r : readers \bullet \neg(\#(issued \triangleright \{r\}) > maxloans)$   
 Logical meaning: "The number of books that any reader borrows must be more than the maximum number of loans allowed."

Figure 3: The modified predicate from the Z-to-English translation task.

The fact that no two participants gave exactly the same English translation of the library specification suggests that people might comprehend even seemingly trivial formal specifications in numerous subtly different ways. This opposes a common belief in the software engineering community: “there is only one way to interpret a formal specification, because of the well defined and unambiguous semantics of the specification language” (Liskov and Berzins, 1979, p. 279). Most participants were able to give consistent English translations of the first three predicates, although the fact that some 25-33% erred in each case may be a cause for some concern. However, none of the participants’ natural language translations of the fourth predicate preserved the meaning of the original Z expression. This particular result may have important implications. Firstly, it suggests that significant properties of formal specifications can actually be lost during interpretation or translation. Secondly, although the fourth predicate could have been simplified to one of several more intuitive forms, the forms of participants’ responses suggests that none had attempted to simplify the complex expression in order to ease their interpretations of the original text. Instead, all appear to have relied upon guesswork based on associations implied by the surrounding context in order to arrive at a plausible, but incorrect, understanding. Specifically, it is thought that all participants obtained the gist of the predicate’s meaning by relating its key linguistic components (i.e. variable identifiers) to their own, misleading preconceptions of real world library systems. This is supported by the fact that all participants gave incorrect responses of the form “No reader may borrow more books than the maximum number of loans allowed.” Perhaps the main conclusion that can be drawn from this task is that people are liable to misinterpret counter-intuitive information when it is presented in the context of a complex formal expression.

Task 2b was designed to test the intuitiveness of conceptual mappings from natural language to formal logic and, simultaneously, to test the common engineering claim that natural language based specifications are prone to ambiguity. Participants were asked to translate the following English requirements description into an appropriate form in the Z notation: “Operation ‘ComputeValue’ outputs the sum of its two inputs squared.” Despite its apparent clarity, this operation’s description is actually open to multiple interpretations because it does not specify whether the two inputs must be squared before or after their addition. Therefore, responses resembling either of the two forms shown in Figure 4 should be considered equally valid despite the fact that they would nearly always generate different solutions for the same inputs. It was predicted that most participants would recognise that more than one form of solution is possible and that they would resolve this dilemma with recourse to knowledge of elementary mathematical principles. In this case, the rules of arithmetic state that multiplication precedes addition wherever there is an absence of parentheses. It was therefore expected that most participants would offer responses resembling the form of Solution A.

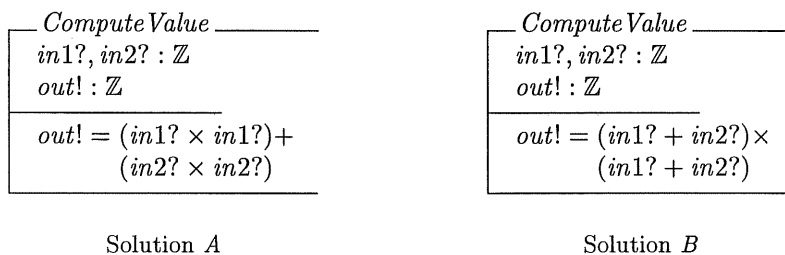


Figure 4: Consistent responses to the English-to-Z translation task.

Responses to Task 2b suggest that, although every participant offered a solution resembling one of the two predicted forms, there was an equally balanced difference of opinion regarding which method was the most appropriate. The fact that half of the participants offered Z implementations resembling Solution *A* and the other half gave responses resembling Solution *B* suggests that natural language based specifications are indeed prone to ambiguous interpretations. Furthermore, the fact that a majority of participants did not give responses resembling the predicted form suggests that software designers should be careful about which aspects of their audience's prior knowledge and mathematical experience are taken for granted. Aside from the method of computation, participants' varied use of the Z notation in Task 2b illustrated several further issues of relevance. It is commonly thought that formal notations constrain the way in which their users write because the number of syntactic symbols and semantic rules that govern them are severely restricted in comparison with, say, those of natural languages. However, participants' responses showed no evidence of this. Despite the apparently limited scope of the English requirements description presented, participants nevertheless offered a wide variety of consistent solutions. In fact, no two responses were exactly the same. This illustrates an important, but often overlooked, issue relating to the production of formal specifications: much is often implied by a requirements description without being explicitly stated in it. Normally, these implicit requirements are implemented by designers according to their own discretions and personal styles of writing. In the case of Task 2b, participants appeared to make implicit but conscious decisions involving at least the following issues: the use of valid and invalid Z notation, the choice of meaningful identifier names, the data types assigned to each variable, the use of parentheses to clarify operator precedence, the ordering of expressions and the use of variables for storing intermediate results. Perhaps above all, the results from the two translation exercises suggest that there rarely exists a clear, unique and intuitive mapping between formal and natural language statements.

“A succinct formulation of a certain property may seem adequate to one; while another will prefer a more verbose exposition of its consequences. To communicate clearly with the majority of readers, you should, in general, prefer clarity to brevity.”

Gravell (1991).

Gravell's claim, based on an informal “straw poll” of software engineers' opinions, suggests that audiences are more likely to understand clear (i.e. precise) specifications rather than brief (i.e. concise) specifications. Intuitively, it seems fair to assume that a clearly written and more detailed specification, as opposed to a brief and abstract one, would be more clearly understood by a majority of its audience. The main aims of the third task were then to discover participants' writing style preferences and, simultaneously, to test Gravell's claim empirically. Participants were presented with an English description of a required software operation, “The operation ‘Toggle’ exchanges the current status of a switch”, and four different Z implementations: one concise, one verbose, one precise and one imprecise (Figure 5). They were then asked to judge which implementation best describes the operation's behaviour and to justify their choices appropriately.

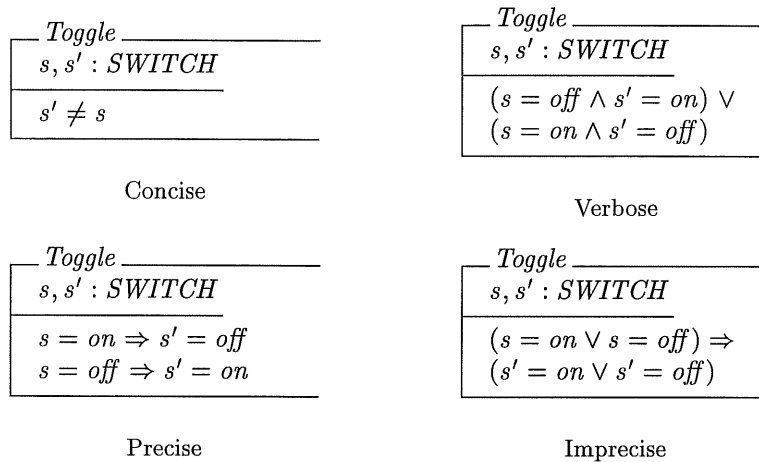


Figure 5: The four styles of formal specification.

Responses to Task 3 indicated that participants' preferences were equally divided amongst three of the four different styles of specification presented, with one third selecting each of the concise, verbose and precise styles. None appeared to favour the imprecise style. This result contradicts Gravell's claim that a majority of readers find precise specifications in particular the easiest to understand and suggests that, whilst precision might be highly desirable, it should not necessarily be the most important factor when writing a formal specification. In fact, the observed responses suggest strong correlations between participants' ages, levels of experience and their style preferences. Whilst the youngest and least experienced tended to choose the concise style, the oldest and most experienced appeared to prefer the precise style. Obviously, the phrase "best describes", used in the task's prompt for responses, is vague and might have had different meanings for different participants. Indeed, the forms of justification offered in support of their choices indicated that the criteria used to discriminate between the four styles did vary. Participants appeared to take into account a combination of linguistic factors before making their selections, but the most influential factors appeared to include: clarity, explicitness, intuitiveness, conciseness, and whether the style of writing was suited the type of application under specification.

"In a spoken situation, the audience is given, but in writing, an audience must be imagined and is to some extent chosen. To write technically is to choose a learned audience."

Turner (1986, p. 190).

The results of the third task appear to have implications for the styles of writing employed by software designers. The extent to which a particular writing style coincides with a reader's natural form of interpretation might depend upon numerous independent variables which add further implicit meaning to what is said explicitly. These include: the type of application being specified, the surrounding linguistic context, the reader's prior knowledge and their language expertise. It might be argued that the ideal level of abstraction that one could use in a specification would take into account its audience's prior knowledge and language expertise. However, in reality, specifications are typically aimed at different readers with differing backgrounds. It is obviously impractical for designers to write several versions, each one aimed at a particular group with a certain level of expertise - for example, designers, programmers, managers and customers. This might explain why precision



is rarely compromised in practice and designers employ a large degree of explicit detail. Whether this principle should be applied in all cases is debatable. Considerate designers writing for novice language users might aim to specify the maximum amount of detail clearly so as to leave nothing to chance, using only the simplest of a notation's constructs. In contrast, considerate designers writing for a learned audience might aim to specify the minimum detail necessary by freely using the full range of a notation's constructs, leaving readers to infer for themselves the other, implicit properties of system functionality. In the former case, this might enable all of a document's potential audience to comprehend, without relying upon readers' knowledge of the notation's more complex features, but at the expense of expert readers finding the document more laborious to read than others. In the latter case, designers would rely entirely upon their audience's expert knowledge of the notation. In this case, there is always the danger that novice readers will not be able to comprehend parts of the specification and will accept the first plausible meaning that appeals to their intuitions, as exemplified by participants' responses to Task 2a. This might be dangerous because readers could then use their inaccurate interpretations as a false basis from which to make incorrect judgements.

An investigation conducted by Evans (1977) aimed to determine whether the linguistic form in which arguments are presented and the presence or absence of negative components would affect the rates at which reasoners drew specific forms of logical inference about conditional statements in natural language. His results suggest that the rates at which people draw successful inferences and succumb to classical fallacies when reasoning about conditional syllogisms can in fact be lowered or raised significantly by manipulating these two independent variables. The purpose of the fourth task was then to determine whether presenting conditional syllogisms in a formal context would affect the rates at which people successfully drew the same valid inferences or succumbed to the same kinds of reasoning fallacies. For each of the fourth task's parts, participants were presented with two premisses in the form of Z predicate expressions: one conditional and one equivalence. Participants were asked to specify what followed from the two premisses by selecting one of four given conclusions. The five forms of inference that participants were required to draw are summarised in Figure 6.

Task 4a (MP): ( <i>shape = circle</i> ) $\Rightarrow$ ( <i>colour = blue</i> ) <i>shape = circle</i> Therefore, <i>colour = blue</i>	Task 4b (affirmative MT): ( <i>shape = circle</i> ) $\Rightarrow$ ( <i>colour = blue</i> ) <i>colour = red</i> Therefore, <i>shape <math>\neq</math> circle</i>
Task 4c (DA fallacy avoided): ( <i>shape = triangle</i> ) $\Rightarrow$ ( <i>colour = red</i> ) <i>shape = square</i> Therefore, nothing follows	Task 4d (AC fallacy avoided): ( <i>shape = square</i> ) $\Rightarrow$ ( <i>colour = green</i> ) <i>colour = green</i> Therefore, nothing follows
Task 4e (negative MT): $\neg$ ( <i>shape = circle</i> ) $\Rightarrow$ ( <i>colour = blue</i> ) <i>colour <math>\neq</math> blue</i> Therefore, <i>shape = circle</i>	

Figure 6: The five conditional inference tasks.

Results from the fourth task provided some insight into the difficulties that people experience when attempting to draw different types of inference and their proneness to classical fallacies when reasoning about conditional syllogisms expressed in formal logic. The fact that every participant responded correctly to Task 4a, in conjunction with the equal success rate observed for the equivalent in-

ference in Evans' natural language based study, suggests that people are generally adept at drawing MP inferences independently of their linguistic context. However, only one third of participants appeared to draw the formal logic based MT inference, in comparison with three quarters drawing the corresponding natural language inference. This suggests that people find it more difficult to reason logically about explicit conditionals in formal logic than implicit conditionals in natural language. It also underlines the relative difficulty that people experience in drawing the MT form of inference, in comparison with the MP form, and might begin to explain the same participants' poor performance on the standard and formalised Wason selection tasks where it is necessary to make an MT inference in order to evaluate the  $\neg q$  case as being relevant. In Evans' study, the presence of a negative operator in the first premiss of a conditional argument's major premiss appears to have a detrimental effect on participants' ability to make MT inferences. The equal success rate observed for Tasks 4b and 4e suggest that participants were not at all distracted by the presence or absence of the negative operator in the formalised study. However, the fact that only one third of participants made the correct responses in each case suggests that participants experienced severe difficulties in drawing both the simple and more complex MT inferences in formal logic. The results from Tasks 4c and 4d suggests a significant difference in the proneness of reasoners to classical fallacies when reasoning about arguments expressed in natural language and formal logic. Participants' responses to these tasks suggest that only around 8% denied the antecedent whilst 17% affirmed the consequent, as compared with respective scores of 69% and 75% in Evans' natural language based study. These vast differences in success rates between the two studies suggest that people may be less prone to commit these two fallacies when reasoning about abstract, formal logic.

In conclusion, results from the first study appear to have important consequences for the future practice of formal software specification. The findings from Tasks 1 and 4 in particular seem to emphasise that people frequently depart from logical rules when reasoning about formal specifications, even when such rules are well known to them. Such results appear to contradict the views of those proponents of mental logic theory who argue that the human mind contains inference rules similar to those found in formal systems of logic and that people who have acquired a high degree of deductive competence would, under ideal conditions, always employ the correct rule at the correct time in order to derive a logically valid conclusion (Inhelder and Piaget, 1958; Rips, 1994). Despite the logical backgrounds of those people who participated in the first experiment, the observed results suggest that their reasoning had frequently strayed from fundamental principles of logical deduction. Results of the first translation task suggest that one particular time at which a reader's faculty for logical reasoning is likely to give way to intuitive guesswork is when counter-intuitive information is encountered in complex formal expressions. The observed responses suggest that, under these circumstances, people are prone to postulate possible meanings and accept the most plausible based on the surrounding context and their prior knowledge of similar software applications. In reality, people are thus liable to interpret formal specifications according to their own heuristic methods which can give rise to interpretations that do not necessarily coincide with those prescribed by a notation's underlying formal semantics. If such interpretations are then used as the premisses for making important project decisions, then it is easy to see why poorly written specifications have in the past led to the development of defective systems.

In the software engineering community, the personal characteristics of designers are currently assumed to have little influence during the production of a formal specification. Yet, participants' responses to the second translation task indicate that, in practice, rarely do any two designers arrive at exactly the same formal specification, even when this is based on the same set of trivial requirements. Thus, the

reality is that formal specification is far from being a completely automated process and that, despite having much more restricted grammars than natural languages, formal notations are still powerful enough to allow designers to exercise their own discretions, creativity and freedom of expression. Perhaps more importantly, this translation task demonstrated that the production of a formal specification is frequently guided by subjective human judgement and informal or undefined processes. It is therefore frequently prone to human error.

The results of the first experiment are significant from a psychological perspective because they suggest that many of the erroneous inferences that people make about implicit logic in natural language also occur when reasoning about explicit logic in formal specifications. However, its results are also significant from a computing science perspective because they suggest that some of the software engineering community's widely held beliefs about formal methods might, in fact, be misconceptions.<sup>1</sup>

### 3.0 Experiments In Progress

#### 3.1 Experiment 2: Reasoning About Conditional Syllogisms in Different Linguistic Contexts

For the purposes of interpreting reasoning data, Evans (1972a) proposes a distinction between factors relating to the interpretation of an argument's premisses and those relating to the reasoning operations required by the inferential task. His later study (Evans, 1977) sought to demonstrate differences between such interpretational and operational factors by manipulating two linguistic variables and observing the effects on people's conditional reasoning performance. Evans argues that manipulating the presence or absence of negative components in conditional rules can reveal operational differences, whilst manipulating the underlying form of logically equivalent "*if p then q*" and "*p only if q*" sentences can reveal interpretational differences. The results of his study suggest that the rates at which people infer correctly or succumb to classical fallacies when reasoning about conditional rules expressed in natural language can indeed be raised or lowered significantly simply by changing these two variables.

Following on from Evans' study, the project's second experiment focuses on people's abilities to draw logically valid inferences and their proneness to commit fallacies when reasoning about conditional rules in formal logic. In a similar manner to Evans' two conditional forms it is hypothesised that, although the formal proposition " $p \Rightarrow q$ " and the natural language sentence "*if p then q*" might be logically equivalent, they are not psychologically equivalent, and interpretational differences between them will affect reasoning performance. Following on from the results of previous psychological studies which suggest that thematic content can facilitate sound reasoning for natural language based conditionals (Cheng and Holyoak, 1985; Griggs and Cox, 1982), it is hypothesised that further interpretational differences will be revealed by monitoring the effects of varying the levels of thematic content in formal specifications. Furthermore, it is predicted that varying the polarity of logical components will reveal operational differences in the methods that people use to reason about conditional arguments in both natural language and formal logic.

The main aims of the second experiment are then to test whether people employ the same or similar cognitive processes when reasoning about logically equivalent conditional rules expressed in different languages and, furthermore, whether reasoning is facilitated by abstract or thematic content. Sixty computer scientists

---

<sup>1</sup>A detailed account of the first experiment was written in the form of a psychology journal paper and submitted as a technical report (Vinter et al., 1996a). Its main findings have also been summarised and oriented towards a computing science audience (Vinter et al., 1996b).

from various academic institutions across the United Kingdom are participating in the between groups study. These are divided equally amongst three experimental groups: abstract natural language (ANL), abstract formal logic (AFL) and thematic formal logic (TFL). Task sheets are completed anonymously and then mailed back to the experimenter. All groups are asked to provide the following biographical details: occupation, age, course, division, year of study, and details of any system of mathematical logic studied beforehand. The two formal logic groups are also asked to provide the following additional information: number of years' Z experience, a list of known formal notations, and a subjective rating of their Z expertise (novice, proficient or expert).

The grammatical frameworks provided by the English language and the Z notation (Spivey, 1992) are used for expressing the natural language and formal logic based tasks respectively. In a similar vein to Evans' original purely natural language based study, the four specific types of inference under scrutiny are: MP, MT, DA and AC. Whilst MP and MT are both valid forms of deductive inference, it should be noted that DA and AC are fallacious forms in which nothing can be concluded logically. Each task requires participants to draw a logical inference from the premisses of a conditional syllogism and then to select the one conclusion that follows logically from four choices. Table 1 illustrates the underlying logical forms of the sixteen tasks.

TABLE 1  
A summary of the valid and fallacious conditional inferences in their abstract forms

Form	MP	MT	DA	AC
1	if $p$ then $q$ , $p$ $\therefore q$	if $p$ then $q$ , $\neg q$ $\therefore \neg p$	if $p$ then $q$ , $\neg p$ $\therefore \neg q$	if $p$ then $q$ , $q$ $\therefore p$
2	if $p$ then $\neg q$ , $p$ $\therefore \neg q$	if $p$ then $\neg q$ , $q$ $\therefore \neg p$	if $p$ then $\neg q$ , $\neg p$ $\therefore q$	if $p$ then $\neg q$ , $\neg q$ $\therefore p$
3	if $\neg p$ then $q$ , $\neg p$ $\therefore q$	if $\neg p$ then $q$ , $\neg q$ $\therefore p$	if $\neg p$ then $q$ , $p$ $\therefore \neg q$	if $\neg p$ then $q$ , $q$ $\therefore \neg p$
4	if $\neg p$ then $\neg q$ , $\neg p$ $\therefore \neg q$	if $\neg p$ then $\neg q$ , $q$ $\therefore p$	if $\neg p$ then $\neg q$ , $p$ $\therefore q$	if $\neg p$ then $\neg q$ , $\neg q$ $\therefore \neg p$

In order to minimise the influence of thematic content, all of the tasks presented to the two abstract groups are set in the context of colours and shapes scenarios. Whereas, in order to encourage the retrieval of relevant knowledge from participants' memories, a variety of realistic scenarios are used for the thematic tasks. These scenarios include: a security door system, a printer manager, a football identity card system, and a student registration system. The experimental tasks presented to the ANL, AFL and TFL groups are exemplified by the AC-2 inference shown in Figures 7, 8 and 9 respectively.

If the shape is a circle then the colour is not blue.  
The colour is not blue.  
Based on the above description, what can you say about shape?

(A) the shape is not a rectangle      (C) the shape is not a circle  
(B) the shape is a circle              (D) nothing

Figure 7: The abstract natural language based AC-2 task.

If  $colour' \neq blue$  after its execution, what can you say about the value of  $shape$  before operation *SetColour* has executed?

<i>SetColour</i>	
$\Delta ShapeAndColour$	
$(shape = circle) \Rightarrow (colour' \neq blue)$	
$shape' = shape$	
	(A) $shape \neq rectangle$
	(B) $shape = circle$
	(C) $shape \neq circle$
	(D) Nothing

Figure 8: The abstract formal logic based AC-2 task.

If  $\neg(reactor\_status! = Ok)$  after its execution, what can you say about  $coolertemp$  before operation *ReactorTempCheck* has executed?

<i>ReactorTempCheck</i>	
$\exists NuclearPlantStatus$	
$reactor\_status! : Report$	
$coolertemp > Maxtemp \Rightarrow \neg(reactor\_status! = Ok)$	
	(A) $coolertemp \leq Maxtemp$
	(B) $coolertemp > Maxtemp$
	(C) $coolertemp > Mintemp$
	(D) Nothing

Figure 9: The thematic formal logic based AC-2 task.

Following each experimental task, participants are asked to give a subjective rating of the extent to which they believe their response is correct. This is achieved by ticking an appropriate box, as follows:

Confidence rating:  Not confident  Guess  Confident

The second experiment is approaching completion and some preliminary results have already been summarised. Overall scores for the three groups suggest that participants are adhering to logical principles more often than not, with the TFL group performing best and the ANL group performing worst. This is perhaps reflected by participants' confidence ratings, which suggest that the TFL group is most confident and the ANL group is least confident about the correctness of their responses. Clear trends are becoming evident in the ways in which participants' are responding to the four main types of inference: MP, MT, DA, and AC. The observed results suggest that people are generally adept at drawing the MP inference irrespective of its surrounding linguistic context, but that they are much more likely to draw valid MT inferences in formal logic than in natural language. Although a high rate of participants have correctly deduced that nothing follows logically from a DA inference in all three groups, participants from the TFL group in particular appear much less prone to denying the antecedent. Finally, response patterns suggest that most participants are failing to recognise that nothing follows logically from an AC inference, although the natural language group does seem to affirm the consequent less frequently than the other two formal logic groups.

Overall, the response patterns observed thus far suggest that Evans' claims regarding the manipulation of those same independent variables which uncovered interpretational and operational factors in natural language are equally valid for formal logic. The performance of one way chi-square tests on the results of each of the sixteen types of conditional inference suggest that the underlying form of arguments and the polarity of logical components do indeed have a significant effect on participants' reasoning performance in each of the three linguistic domains. The response patterns suggest that participants are employing the same or similar cognitive processes when drawing formal logic and natural language based inferences.

But, despite substantial differences in the three groups' overall scores, performance of a two way chi-square test comparing the results of the AFL and ANL groups suggests that, although the concepts of " $p \Rightarrow q$ " and "*if p then q*" might differ psychologically, the ways in which people reason about them does not differ significantly. Furthermore, although the use of thematic content appears to enhance formal reasoning slightly, the performance of a second two way chi-square test comparing the AFL and TFL groups' responses suggests that this difference in performance is not significant statistically.

In view of the fact that formal notations are gradually gaining favour over natural language based notations, the preliminary results of the second investigation appear to have far-reaching implications for the software engineering community because they suggest that, under certain circumstances, the use of formal logic based descriptions can lead to slightly enhanced reasoning performance. The fact that the TFL group has thus far fared slightly better than the other two groups suggests that the use of formal logic and thematic content can increase the likelihood an audience will interpret and reason about a conditional expression in the manner intended by its author. However, the fact that reasoners do still frequently err in all three linguistic contexts should be a cause for concern.<sup>2</sup>

### 3.2 Experiment 3: Reasoning About Disjunctive and Conjunctive Syllogisms in Formal Logic

The connective "*or*" is used liberally and indiscriminately in everyday language to join and express choice between two statements, and many writers appear to take for granted the complex linguistic rules and hidden conventions that govern its use. Hurford (1974) argues that *or* is misused wherever one sentence entails the other. For example, sentences such as "John is British or American" would be legal, whereas "John is British or a Londoner" should be avoided. This is not to suggest that the two conjoined sentences must be completely unrelated as in, for example, "The car is red or London is the capital of England." Indeed, the overwhelming consensus is that any two conjoined sentences in the English language must share a common topic (Lakoff, 1971; Fillenbaum, 1974). But aside from its frequent misuse by writers, *or* also gives rise to ambiguous interpretations in readers. For many years, psychologists and linguists have tried to establish the precise conditions under which readers are obliged to draw inclusive or exclusive interpretations. For example, "Do you want tea or coffee?" normally requires an exclusive interpretation, whereas "Do you want milk or sugar?" suggests an inclusive interpretation. The views of psychologists and linguists are extremely varied, with some arguing that the basic *or* in English is generally inclusive (Pelletier, 1977), some claiming that it is nearly always exclusive (Lakoff, 1971), and some arguing that a correct interpretation depends upon factors such as the context of the sentence and the form of the words used (Hurford, 1974; Newstead and Griggs, 1983). Figure 10 shows that the conditions under which this ambiguity arises is when both disjuncts are true: under an inclusive interpretation the whole sentence would be true, whereas under an exclusive interpretation the sentence would be false.

P	Q	Inclusive Or	Exclusive Or
F	F	F	F
F	T	T	T
T	F	T	T
T	T	T	F

Figure 10: Logical truth-tables for inclusive and exclusive *Or*.

<sup>2</sup>At the time of writing, the results of the second experiment are being gathered and summarised (Vinter et al., 1996c).

Previous studies of disjunctive reasoning have shown that, where this form of ambiguity exists, people generally have a preference for exclusive interpretations although the strength of this preference has been found to vary according to the context in which the disjunctive is presented (Newstead and Griggs, 1983; Newstead et al., 1984). Developmental studies point to evidence which suggests that people tend to begin childhood with a strong preference for inclusive interpretations and then gradually develop a preference for exclusive interpretations with age (Sternberg, 1979; Braine and Rumin, 1981). However, one factor that might confound the results of such studies is the possibility that young children respond to disjunctive statements as if they were conjunctions which gives the false impression that they are adopting inclusive interpretations. Wason and Johnson-Laird (1969) observed a similar phenomenon when adult participants appeared to confuse the principles of disjunction and conjunction. However, once the correct interpretation has been adopted, people generally find it easier to reason about exclusive disjunctives rather than inclusive disjunctives (Newstead et al., 1984; Roberge, 1977; 1978). Two explanations for this apparent difference in complexity are offered. First, it might be that people are more adept at reasoning with exclusive disjunctives simply because this is the more common form in everyday language. Second, the difference might be explained by the fact that exclusive disjunctives lead to symmetrical inferences: by knowing the truth-value of one disjunct the truth-value of the other can be determined. This contrasts with inclusive disjunctives, where simply knowing the truth-value of one disjunct is an insufficient means for determining the truth-value of the other. Nevertheless, it would appear that reasoning performance is significantly enhanced when it is clarified to reasoners which interpretation is to be drawn (Newstead and Griggs, 1983). This has been achieved in previous studies through the use of qualifying expressions in the disjunctive rules: “*p or q (or both)*” for inclusive disjunction, “*p or q (but not both)*” for exclusive disjunction.

Despite their frequent occurrence in everyday language, mathematical logic has developed an almost universal bias against exclusive disjunction and tends to favour the inclusive form alone. The extent of this bias is typified by standard propositional logic. Although the propositional operator “ $\vee$ ” was derived from the Latin term for inclusive disjunction, *vel*, it might appear odd that there was no corresponding operator derived from its term for exclusive disjunction, *aut*, in a similar fashion. Newstead and Griggs (1983) offer two possible explanations for this bias towards inclusive *or*. Firstly, it is advantageous from a parsimonious perspective because all other logical operations can be defined in terms of it and negation. Secondly, the inclusive operator neatly complements the set union operator “ $\cup$ ” because both refer to either one of two propositions or sets, and possibly both.<sup>3</sup>

$p \text{ or } q$	$p \text{ or } q$
$\text{not } p$	$p$
-----	-----
<i>Or denial</i>	<i>Or affirmation</i>
$q$	$\text{not } q$

Figure 11: Denial and affirmation inferences for *Or*.

The first argument in Figure 11 is often referred to as a “denial inference” because the minor premiss explicitly denies a component from the major premiss, resulting in the affirmation of the other. It is logically valid under either an inclusive or an exclusive interpretation of the disjunctive rule in the major premiss. The second inference is termed an “affirmation inference” because the minor premiss affirms one of the propositions in the major premiss, resulting in the denial of the other. However, it is valid only under an exclusive interpretation of the disjunctive

<sup>3</sup>Logicians often find it necessary to invent non-standard symbols such as Diller (1994, p. 12), who defines the “ $\|$ ” symbol to express exclusive disjunction.

rule and is indeterminate under an inclusive interpretation (since both disjuncts might be true). A summary of several studies conducted by Evans et al. (1993, p. 143-145) suggests that the denial inference is made correctly around 84% and 80% of the time for exclusive and inclusive disjunctives, respectively. This summary also suggests that around 83% draw the affirmation inference correctly for exclusive disjunctives, but that around 36% still fallaciously draw it for inclusive disjunctives where it constitutes a logical error. It is often postulated that people's poor performance on the affirmation task for inclusive disjunctives might be explained by the fact that the correct conclusion is indeterminate rather than true or false.

"It is tempting to attribute the difficulty of disjunctive concepts to the fact that the word 'or' is so ambiguous in the English language. . . . The crucial test of the hypothesis is to compare concept formation across different languages. One would expect that in languages where the word for disjunction is less ambiguous, performance on disjunctive concepts should be better."

Newstead and Griggs (1983).

It is hard to envisage a language that could be much more precisely defined than one whose syntactic and semantic rules are defined entirely by mathematics. These are exactly the type of languages that are currently being used within the software engineering community for the purposes of software and hardware specification. In formal logic, disjunctives are not constrained by the same types of linguistic rules and conventions that govern the use of *or* in the English language. For example, formal statements can be used to connect any two disjuncts regardless of whether they share a common topic. Furthermore, in certain branches of formal logic such as the standard propositional calculus, the concept of exclusive disjunction is undefined. In this light, one might expect that formal logicians do not encounter the kinds of linguistic problems experienced by those people who use disjunctives in natural language. However, the fact that several studies have found that people are inherently biased towards exclusive interpretations in everyday language gives cause to suggest that people might still frequently reason according to exclusive principles when they encounter inclusive disjunctives in formal logic. If this is case, then are formal logicians being asked to abandon natural intuition and reason according to formally defined rules alone? These are the kinds of issue that the project's third experiment is aimed at illuminating.

In contrast with the connectives *if* and *or*, there has been very little psychological research aimed at investigating the ways in which people reason about *and*. One possible explanation for this is that there may be less complex linguistic rules and conventions governing the use of conjunctives and the possibility that people may be less prone to error and bias during conjunctive reasoning. However, Lakoff (1971) states that the principles governing the use of disjunction and conjunction in the English language are actually quite similar because both require a common topic between two conjuncts and this may either be overtly present or derivable by presupposition and deduction. He points to the existence of a hierarchy of sentences conjoined by *and* with varying degrees of relationship between their common topics. At the top of this hierarchy are sentences like "John eats apples and John eats pears", where the meaning of one of the conjoined sentences complements the meaning of the other. At the bottom we might find sentences whose common topics have a negative relationship such as "John is a strict vegetarian and he eats lots of meat", where one of the conjoined sentences directly opposes the meaning of the other. However, there are obvious differences between conjunction and disjunction. For example, Lakoff points to the fact that the successful interpretation of a conjunctive sentence relies upon the presupposition of the first conjunct in order to facilitate understanding of the second. This contrasts with disjunction, where



the truth of the first disjunct is never presupposed, although its negation might be presupposed in order for the second disjunct to be considered true.

“In short, reasoning is nothing more than the propositional calculus itself. Although, in the subject’s thought, this calculus is linked to current speech patterns, it can be expressed symbolically in terms of the algebra of propositional logic.”

Inhelder and Piaget (1958, p. 305).

Mental logic theory claims that logic itself is a normative theory of deductive competence and that reasoning is guided by inference rules similar to those found in formal systems of logic. These include rules for propositional connectives and quantifiers such as: *and*, *or*, *if*, *not*, *all* and *every*. In general, these theorists assume that people who have acquired a high degree of deductive competence would, under ideal conditions, always employ the correct rule at the correct time to enable them to derive the correct conclusion (Inhelder and Piaget, 1958; Rips, 1994). If this theory were true, then not only should the rules for conjunctive introduction and elimination (Figure 12) be part of every person’s mental machinery, but they should be two of the more basic rules that formal logicians command. The third experiment aims to test this hypothesis under a formal context.

$$\begin{array}{ccc}
 \begin{array}{c} p \\ q \\ \hline p \text{ and } q \end{array} & \text{And intro} & \begin{array}{c} p \text{ and } q \\ \hline p \end{array} & \text{And elim 1} & \begin{array}{c} p \text{ and } q \\ \hline q \end{array} & \text{And elim 2}
 \end{array}$$

Figure 12: Logical rules of inference for *And*.

Probability theory states that the likelihood of a conjunction, *A and B*, cannot exceed the likelihood of one of its constituent outcomes, *A* or *B*, because the possibility set of the conjunction is included in the extension of its constituents. A series of studies performed by Tversky and Kahneman (1983) sought to uncover conflicts between logic and intuition when reasoning about conjunctions. Specifically, they aimed to test whether people’s adherence to this “conjunctive fallacy” persisted across a variety of different contexts. Their results suggest that violation is likely whenever reasoners depart from the principles of logic and start adhering to intuitive heuristics, such as those of representativity and availability. The representativeness heuristic states that people are likely to judge the overall conjunction as more representative of an particular category than its individual constituents. The availability heuristic states that instances of a more inclusive category are easier to imagine and retrieve than those of a specific category. Responses from the first experiment (Vinter et al., 1996a) suggest that, when complex expressions are encountered in formal specifications, reasoners are liable to abandon logical principles and rely upon guesswork based on associations implied by the surrounding context and prior knowledge in order to arrive at plausible judgements. It is an aim of the third study to test whether the representativity and availability heuristics observed by Tversky and Kahneman are also adopted by formal logicians and, hence, whether the conjunctive fallacy transfers into the domain of formal reasoning.

$$\begin{array}{ccc}
 \begin{array}{c} \text{not } (p \text{ or } q) \\ \hline \text{not } p \text{ and not } q \end{array} & \text{Not over Or} & \begin{array}{c} \text{not } (p \text{ and } q) \\ \hline \text{not } p \text{ or not } q \end{array} & \text{Not over And}
 \end{array}$$

Figure 13: De Morgan’s laws.

If mental logic theory reflected reality then it would be reasonable to expect that reasoners who have acquired a high degree of deductive competence would have acquired some of the more sophisticated rules of inference and might even be capable of performing multiple stages of reasoning. Figure 13 shows two of de Morgan's laws which are commonly used in logic to convert disjunctions into conjunctions and vice versa (Diller, 1994, p. 22; Lemmon, 1993, p. 62). Although they are perhaps not used as often as propositional logic's more basic rules, such as the introduction and elimination rules for conditionals, disjunctives and conjunctives, most formal logicians will almost certainly have used them beforehand. However, the question of whether they see their relevance to a given situation is another matter. Thus, in order to test participants' competence in performing multi-stage disjunctive and conjunctive reasoning, a series of tasks was devised for the third experiment which required the application of one of de Morgan's laws, followed by the application of a simple disjunctive or conjunctive elimination rule. If, as several developmental studies have shown, deductive competence increases with age, then one would expect only those expert reasoners with extensive backgrounds in formal logic and many years' experience to perform well on these tasks.

Forty computer scientists from various academic institutions across the United Kingdom are participating in the third study which has a between groups design: two sets of twenty eight tasks containing abstract and thematic content are presented to two corresponding linguistic groups. All reasoning tasks are syllogistic in nature and based around disjunctive and conjunctive inferences. The disjunctive elimination (DE) tasks involve either the denial of a component (DC) or the affirmation of a component (AC) from a major premiss, by the minor premiss. The polarity of components in the major premiss and the position of the component affirmed or denied by the minor premiss are varied systematically. Table 2 illustrates the underlying forms of the sixteen disjunctive tasks presented to participants.

TABLE 2  
A summary of the disjunctive denial and affirmation inferences

Major premiss	DE-DC	DE-AC
AA	$p \vee q, \neg p \therefore q$ $p \vee q, \neg q \therefore p$	$p \vee q, p \therefore \neg q$ $p \vee q, q \therefore \neg p$
AN	$p \vee \neg q, \neg p \therefore \neg q$ $p \vee \neg q, q \therefore p$	$p \vee \neg q, p \therefore q$ $p \vee \neg q, \neg q \therefore \neg p$
NA	$\neg p \vee q, p \therefore q$ $\neg p \vee q, \neg q \therefore \neg p$	$\neg p \vee q, \neg p \therefore \neg q$ $\neg p \vee q, q \therefore p$
NN	$\neg p \vee \neg q, p \therefore \neg q$ $\neg p \vee \neg q, q \therefore \neg p$	$\neg p \vee \neg q, \neg p \therefore q$ $\neg p \vee \neg q, \neg q \therefore p$

The conjunctive reasoning tasks involve either the fallacious introduction (CI) or the valid elimination (CE) of logical components. The polarity of components in the arguments' premisses and the types of component introduced or eliminated are systematically varied. The logical inferences underlying the eight conjunctive reasoning tasks are summarised in their abstract forms in Table 3.

TABLE 3

A summary of the conjunctive elimination and introduction inferences

Form	CE	CI
1	$p \wedge q \therefore p$	$p \therefore p \wedge q$
2	$p \wedge \neg q \therefore \neg q$	$q \therefore p \wedge q$
3	$\neg p \wedge q \therefore \neg p$	$\neg p \therefore \neg p \wedge q$
4	$\neg p \wedge \neg q \therefore \neg q$	$\neg q \therefore p \wedge \neg q$

The conjunctive and disjunctive reasoning tasks requiring the use of de Morgan's laws (Diller, 1994, p. 22; Lemmon, 1993, p. 62; Rips, 1994, p. 113-114) are potentially the most difficult of all the inferences that participants are asked to draw because each involves the application of two rules of inference: either de Morgan's over disjunction followed by conjunction elimination (DMD-CE), or de Morgan's over conjunction followed by disjunctive elimination (DMC-DE). In view of the anticipated difficulties that participants will experience in drawing these inferences, the polarity of the two components in their premisses are being held constant and only the type of component eliminated is varied. The underlying structures of these inferences are illustrated in Table 4.

TABLE 4

The disjunctive and conjunctive inferences requiring de Morgan's law

Form	DMD-CE	DMC-DE
1	$\neg(p \vee q) \therefore \neg p$	$\neg(p \wedge q), p \therefore \neg q$
2	$\neg(p \vee q) \therefore \neg q$	$\neg(p \wedge q), q \therefore \neg p$

In order to minimise the possibility of interference from prior knowledge, colours and shapes scenarios are employed throughout the abstract versions of the tasks. In contrast, the formal specifications of imaginary but realistic scenarios are presented to members of the thematic group. These scenarios include a missile guidance system, a nuclear reactor status check, a live event's television coverage, a telephone network, and a hotel reservation system. The experiment's materials are exemplified by the abstract and thematic versions of the DE-AA-DC inference task shown in Figures 14 and 15, respectively.

If  $\neg(\text{colour!} = \text{white})$  what can you say about  $\text{shape!}$  in operation *GetShapeColour*?

<i>GetShapeColour</i>	
$\text{shape!} : \text{SHAPE}$	(a) $\text{shape!} = \text{square}$
$\text{colour!} : \text{COLOUR}$	(b) $\text{shape!} = \text{circle}$
	(c) $\neg(\text{shape!} = \text{rectangle})$
$\text{colour!} = \text{white} \vee \neg(\text{shape!} = \text{rectangle})$	(d) Nothing

Figure 14: Abstract version of a denial inference task.

If  $\neg(\text{password!} = \text{Correct})$  after its execution, what can you say about *report!* in operation *AccessSystem*?

<i>AccessSystem</i> <i>password!</i> : <i>Status</i> <i>report!</i> : <i>Report</i>
$\text{password!} = \text{Correct} \vee \neg(\text{report!} = \text{Unauthorised})$

- |  |  |
|--|--|
| (a) <i>report!</i> = <i>Unauthorised</i> | (c) $\neg(\text{report!} = \text{Unauthorised})$ |
| (b) <i>report!</i> = <i>Authorised</i>   | (d) Nothing                                      |

Figure 15: Thematic version of a denial inference task.

Before completing the test, participants are asked to provide brief biographical details including: occupation, age, organisation, course, number of years' Z experience, a list of other formal notations known, a subjective rating of their Z expertise (novice, proficient or expert), and details of any system of formal logic studied beforehand (for example, the propositional or predicate calculus). Following each task, participants are asked to give a subjective rating of the extent to which they believe their responses to be correct. This is achieved by ticking an appropriate box, in the same manner described for Experiment 2.

Owing to the limited number of completed task sheets received thus far, it is too early to point to the existence of clear patterns in participants' responses. However, preliminary response data does suggest that participants are finding the purely disjunctive inferences slightly more difficult than the purely conjunctive inferences. The error rates observed thus far also suggest that reasoners are experiencing severe difficulties in solving those tasks requiring the application of de Morgan's law, with around 50% erring on the DMC-DE tasks and 30% erring on the DMD-CE tasks.<sup>4</sup>

## 4.0 Planned Experiments

In addition to the project's first three investigations, a further three experiments are planned in order to generate the raw data necessary for the intended formulation of specification metrics. Their designs have yet to be finalised at the time of writing.

### 4.1 Experiment 4: Reasoning About Categorical Syllogisms Expressed in Formal Logic

Previous studies have shown that people are prone to various errors and biases when reasoning about quantified statements forming the premisses of syllogistic arguments. Woodworth and Sells (1935) highlight the occurrence of a non-logical "atmosphere effect," Johnson-Laird (1977) and Dickstein (1978a) point to the occurrence of a "figural effect," Dickstein (1978b) demonstrates that reasoners illicitly integrate and convert premisses and fail to consider certain hypothetical possibilities, and Erickson (1974; 1978) suggests that people frequently depart from the set theoretic principles which underly valid syllogistic reasoning. Indeed, Evans et al. (1993, p. 217-240) show that reasoners are liable to err at all three stages of syllogistic reasoning: interpretation, premiss combination, and response generation. All such previous psychological studies have, however, been set in the context of natural language alone. Hence, the project's fourth experiment aims to test whether there is any improvement in people's abilities to reason logically about universally and existentially quantified predicates forming categorical syllogisms in the Z notation.

<sup>4</sup>An analysis of the third study's results is currently under preparation (Vinter et al., 1996d).

## 4.2 Experiment 5: Reasoning About Complex Formal Expressions

Discounting the presence of negative components, the project's first three investigations concentrate on reasoners' abilities to draw inferences about formal predicates containing only a single logical operator ( $\Rightarrow$ ,  $\vee$ , and  $\wedge$ ). These are certainly quite different from the more complicated forms of logical expression that designers tend to encounter in everyday life, each of which may contain a different combination of multiple quantifiers and propositional connectives. As such, the results of these investigations may not be generalisable to the kinds of reasoning that occur in real-life situations. Nor would these results be directly comparable to those of previous empirical studies which focus on reasoning with expressions containing combinations of multiple operators, such as Johnson-Laird's (1969) study of reasoning with multiply-quantified English sentences. Hence, the fifth study will focus on reasoning about predicates containing multiple logical operators in an attempt to identify particular combinations that may give rise to reasoning difficulties.

## 5.0 Acknowledgements

The author would like to thank Martin Loomes and Diana Kornbrot for their help in the production of this report, and the following reviewers for their many helpful comments: Richard Ralley, Jane Simpson and Mark Surman. The author would also like to thank all of the staff and students who have participated in the project's experiments. The research reported here is supported by Grant No. J00429434043 from the Economic and Social Research Council.

## References

- Barston, J.L. (1986). An investigation into belief biases in reasoning. Unpublished PhD thesis, University of Plymouth.
- Bowen, J.P. and Hinchey, M.G. (1994a), *Ten commandments of formal methods*, Technical Report No. 350, University of Cambridge Computer Laboratory, September 1994.
- Braine, M.D.S. and O'Brien, D.P. (1991). A theory of If: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, 98, 182-203.
- Braine, M.D.S. and Rumin, B. (1981). Development of comprehension of "or": Some evidence for a sequence of competencies. *Journal of Experimental Child Psychology*, 31, 46-70.
- Cheng, P.W. and Holyoak, K.J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, 17, 391-416.
- Dickstein, L.S. (1978a). The effect of figure on syllogistic reasoning. *Memory and Cognition*, 6 (1): 76-83.
- Dickstein, L.S. (1978b). Error processes in syllogistic reasoning. *Memory and Cognition*, 5, 537-543.
- Diller, A. (1994). *Z. An Introduction to Formal Methods*. Second Edition, Chichester: John Wiley and Sons.
- Erickson, J.R. (1974). A set analysis theory of behaviour in syllogistic reasoning tasks. In R.L. Solso (Ed.), *Theories of Cognitive Psychology: The Loyola Symposium*. Hillsdale, NJ: LEA.
- Erickson, J.R. (1978). Research on syllogistic reasoning. In R. Revlin and R.E. Mayer (Eds.), *Human Reasoning*, Washington, DC: Winston.
- Evans, J.St.B.T. (1972a). On the problems of interpreting reasoning data: logical and psychological approaches. *Cognition*, 1, 373-384.
- Evans, J.St.B.T. (1972b). Interpretation and matching bias in a reasoning task. *Quarterly Journal of Experimental Psychology*, 24, 193-199.

- Evans, J.St.B.T. (1977). Linguistic factors in reasoning. *Quarterly Journal of Experimental Psychology*, 29, 297-306.
- Evans, J.St.B.T. (1983). Linguistic determinants of bias in conditional reasoning. *Quarterly Journal of Experimental Psychology*, 35A, 635-644.
- Evans, J.St.B.T., Barston, J.L., and Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition*, 11 (3): 295-306.
- Evans, J.St.B.T., Newstead, S.E., and Byrne, R.M.J. (1993). *Human Reasoning. The Psychology of Deduction*. Hove (UK): Lawrence Erlbaum Associates.
- Fillenbaum, S. (1974). Or: some uses. *Journal of Experimental Psychology*, 103, 913-921.
- Gravell, A. (1991). What is a good formal specification? In J.E. Nicholls (Ed.), *Z User Workshop, Oxford 1990. Proceedings of the Fifth Annual Z User Meeting, Oxford 17-18 December 1990*, Springer-Verlag.
- Griggs, R.A. and Cox, J.R. (1982). The elusive thematic materials effect in the Wason selection task. *British Journal of Psychology*, 73, 407-420.
- Griggs, R.A. and Jackson, S.L. (1990). Instructional effects in Wason's selection task. *British Journal of Psychology*, 81, 197-204.
- Hurford, J.R. (1974). Exclusive or inclusive disjunction. *Foundations of Language*, 11, 409-411.
- Inhelder, B. and Piaget, J. (1958). *The Growth of Logical Thinking From Childhood to Adolescence: An Essay on the Construction of Formal Operational Structures*. New York: Basic Books.
- Johnson-Laird, P.N. (1969). On understanding logically complex sentences. *Quarterly Journal of Experimental Psychology*, 21, 1-13.
- Johnson-Laird, P.N. (1977). Reasoning with quantifiers. In P.N. Johnson-Laird and P.C. Wason (Eds.), *Thinking. Readings in Cognitive Science*, 129-142, Cambridge: Cambridge University Press.
- Johnson-Laird, P.N. and Tridgell, J.M. (1972). When negation is easier than affirmation. *Quarterly Journal of Experimental Psychology*, 24, 87-91.
- Johnson-Laird, P.N. and Wason, P.C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, 1, 134-138.
- Lakoff, R. (1971). If's, and's, and but's about conjunction. In C.J. Fillmore and D.T. Langendoen (Eds.), *Studies in Linguistic Semantics*. New York: Holt, Rinehart and Winston.
- Lemmon, E.J. (1993). *Logic*. Second Edition, London: Chapman and Hall.
- Liskov, B. and Berzins, V. (1979). An appraisal of program specifications. In P. Wegner (Ed.), *Research Directions in Software Technology*, Cambridge, Mass: MIT Press.
- Morgan, J.J.B. and Morton, J.T. (1944). The distortion of syllogistic reasoning produced by personal convictions. *Journal of Social Psychology*, 20, 39-59.
- Newstead, S.E. and Griggs, R.A. (1983). The language and thought of disjunction. In J.St.B.T. Evans (Ed.), *Thinking and Reasoning. Psychological Approaches*, 76-106, London: Routledge and Kegan Paul.
- Newstead, S.E., Griggs, R.A. and Chrostowski, J.J. (1984). Reasoning with realistic disjunctives. *Quarterly Journal of Experimental Psychology*, 36A, 611-627.
- Oakhill, J., Garnham, A., and Johnson-Laird, P.N. (1990). Belief bias effects in syllogistic reasoning. In K.J. Gilhooly, M.T.G. Keane, R.H. Logie, and G. Erdos (Eds.), *Lines of Thinking: Reflections on the Psychology of Thought. Volume 1. Representation, Reasoning, Analogy and Decision Making*, 125-138, John Wiley and Sons.
- Pelletier, F.J. (1977). Or. *Theoretical Linguistics*, 4, 61-74.
- Potter, B., Sinclair, J. and Till, D. (1991). *An Introduction to Formal Specification and Z*. Hemel Hempstead: Prentice-Hall International.

- Rips, L.J. (1994). *The Psychology of Proof: Deductive Reasoning in Human Thinking*. Cambridge, Mass: MIT Press.
- Roberge, J.J. (1977). Effects of content on inclusive disjunction reasoning. *Quarterly Journal of Experimental Psychology*, 29, 669-676.
- Roberge, J.J. (1978). Linguistic and psychometric factors in propositional reasoning. *Quarterly Journal of Experimental Psychology*, 30, 705-716.
- Smith, N.K. (Trans., 1993), *Immanuel Kant's Critique of Pure Reason*. Second Edition, London: Macmillan, 1993.
- Spivey, J.M. (1992). *The Z Notation: A Reference Manual*. Second Edition. Hemel Hempstead, Prentice Hall International.
- Turner, G.W. (1986). *Stylistics*, Harmondsworth: Penguin Books.
- Tversky, A. and Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review*, 90 (4): 293-315.
- Sternberg, R.J. (1979). Developmental patterns in the encoding and comprehension of logical connectives. *Journal of Experimental Child Psychology*, 28, 469-498.
- Vinter, R., Loomes, M., and Kornbrot, D. (1996a). *Reasoning About Formal Software Specifications: An Initial Investigation*, Technical Report No. 249, Division of Computer Science, University of Hertfordshire, March 1996.
- Vinter, R., Loomes, M. and Kornbrot, D. (1996b). *Seven Lesser Known Myths of Formal Methods: Uncovering the Psychology of Formal Specification*, Technical Report No. 250, Division of Computer Science, University of Hertfordshire, April 1996.
- Vinter, R., Loomes, M., and Kornbrot, D. (1996c). *Conditional Reasoning Between Linguistic Contexts: Formal Logic Versus Natural Language*, unfinished technical report, Division of Computer Science, University of Hertfordshire, available from October 1996.
- Vinter, R., Loomes, M., and Kornbrot, D. (1996d). *A Study of Disjunctive and Conjunctive Formal Reasoning*, unfinished technical report, Division of Computer Science, University of Hertfordshire, available from November 1996.
- Wason, P.C. (1966). Reasoning. In B.M. Foss (Ed.), *New Horizons in Psychology. Volume 1*, Reading: Penguin.
- Wason, P.C. and Johnson-Laird, P.N. (1969). Proving a disjunctive rule. *Quarterly Journal of Experimental Psychology*, 21, 14-20.
- Wason, P.C. and Johnson-Laird, P.N. (1972). *Psychology of Reasoning: Structure and Content*. London: Batsford.
- Wason, P.C. and Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, 23, 63-71.
- Woodworth, R.S. and Sells, S.B. (1935). An atmosphere effect in syllogistic reasoning. *Journal of Experimental Psychology*, 18, 451-460.