DEPARTMENT OF COMPUTER SCIENCE

# A Study of Disjunctive and Conjunctive Reasoning in Formal Logic

R J Vinter
M J Loomes
D E Kornbrot

Technical Report No. 298

April 1997

# A Study of Disjunctive and Conjunctive Reasoning in Formal Logic

Rick Vinter and Martin Loomes
*Faculty of Information Sciences*

Diana Kornbrot
*Faculty of Health and Human Sciences*

*University of Hertfordshire, Hatfield, UK* [1]

When reasoning with statements containing logical connectives in everyday discourse, people sometimes employ reasoning strategies that do not comply with the dictates of logic yet are still sufficient for their purpose. Cognitive studies point to the most likely circumstances in which non-logical heuristics are likely to be employed for logical problems expressed in natural language, where they invariably lead to error. This paper describes a study aimed at determining whether trained computer scientists continue to employ non-logical heuristics when they are reasoning about logical statements expressed in a mathematical notation. The study focuses on the ways in which people reason about disjunctive and conjunctive statements. Specifically, it sought to test whether reasoning performance is affected by the polarity of logical operators or by the degree of thematic content presented in problem material. The results suggest that there was only a limited transfer of non-logical processes to the formal domain and that, although reasoning was still far from perfect, the use of a formal notation facilitated logical reasoning for the types of inference scrutinised. The implications of this finding are discussed in relation to the software engineering community where the use of formal logic based notations are gradually gaining increased acceptance.

## Introduction

Historically, in the context of software development, poorly written software specifications have led to the production of defective software systems (Potter et al., 1991). A large proportion of the defects have been attributed to the imprecision and verbosity introduced through the use of conventional natural language based techniques (Gehani, 1986; Ince, 1992). Proponents of formal methods often claim distinct advantages over natural language for the purpose of software specification such as increased precision and concision, the provision of an objective basis for verifying correctness, and the increased levels of insight and confidence gained through the use of a mathematical approach (Bowen, 1988; Liskov and Berzins, 1986; Sommerville, 1992). However, many of these claims are based on isolated case studies and anecdotal evidence from which results cannot easily be generalised. When one recognises the increasingly business- and safety-critical nature of the systems now being developed with formal methods and the increasing severity of failing to detect defects in delivered systems, one begins to appreciate the importance of subjecting such claims to thorough empirical examination.

Imperfect reasoning invariably gives rise to erroneous decisions and, in the context of software engineering, erroneous decisions can lead to the introduction of defects in software systems. During the past three decades, cognitive studies have shown that people are prone to various forms of systematic error and bias when

[1]The authors would like to thank the following reviewers for their help in the production of this paper: Richard Ralley and Jane Simpson. The authors would also like to thank all of the computing staff, students and professionals who took part in the experiment. The research reported here was supported by Grant No. J00429434043 from the Economic and Social Research Council. Requests for reprints of this paper should be sent to Rick Vinter, Faculty of Information Sciences, Hatfield, Hertfordshire, AL10 9AB, UK.

reasoning about natural language statements containing logical operators such as: "if" (Braine and O'Brien, 1991), "and" (Lakoff, 1971), "or" (Newstead et al., 1984), "not" (Johnson-Laird and Tridgell, 1972), "some" and "all" (Johnson-Laird, 1977). Despite obvious syntactic differences, most formal notations contain logical operators with roughly equivalent semantical definitions as these same natural language constructs: $\Rightarrow$, $\wedge$, $\vee$, $\neg$, $\exists$ and $\forall$. Of course, one of the main questions that the software community must ask is: do the same non-logical errors and biases that people exhibit when reasoning about natural language also occur when software developers are reasoning about the logically equivalent statements in formal specifications?

The present study was conducted as part of a series of investigations aimed at determining whether the same non-logical heuristics that people exhibit when reasoning about logical statements in natural language transfer over into the formal domain. Results from an initial study (Loomes and Vinter, 1997; Vinter et al., 1996) suggest that, under certain circumstances, even trained computer scientists are liable to abandon logical principles during formal reasoning in favour of intuition and guesswork in order to arrive at plausible, rather than logically valid, decisions. Results from a later study (Vinter et al., 1997) suggest that the rates at which trained computer scientists draw valid inferences and succumb to logical fallacies when reasoning about formal specifications containing conditional statements varies according to the presence or absence of specific linguistic properties. The present study aims to test whether similar kinds of linguistic variable are influential when trained computing personnel are reasoning about formal specifications containing disjunctive and conjunctive statements.

## Error and Bias in Disjunctive Reasoning

The logical connective "or" tends to be freely used in everyday language to join and express choice between two statements. However, many writers seem to take for granted the complex linguistic rules and hidden conventions that govern its use. For example, Hurford (1974) argues that a disjunction is misused wherever one sentence entails the other. Hence, sentences such as "John is British or American" should be considered legal, whereas "John is British or a Londoner" should be avoided. But this is not to suggest that the two conjoined sentences should 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).

Aside from its common misuse by writers, disjunctive statements can incite ambiguous interpretations in readers. For decades, psychologists and linguists have tried to establish the precise conditions under which readers are obliged to draw inclusive or exclusive interpretations. Besides what is written explicitly, extra linguistic factors such as context, register and intonation can provide additional clues to a speaker's intended meaning and, thus, help to resolve ambiguities (Turner, 1986). Normally, for example, when we are offered refreshment, the question "Do you want tea or coffee?" requires an exclusive interpretation, whereas "Do you want milk or sugar?" requires an inclusive interpretation, and we are able to determine this with the help of contextual clues. Historically, the resolution of disjunctive ambiguities in everyday language has been strongly debated, 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 others arguing that the correct interpretation always depends upon linguistic factors such as the context of the sentence and the form of the words used (Hurford, 1974; Newstead and Griggs, 1983). Figure 1 shows that the only condition 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 1: Logical truth-tables for inclusive and exclusive disjunction

Previous studies of disjunctive reasoning suggest that, where this form of ambiguity exists, people generally prefer to draw 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 suggest that people tend to begin childhood with a strong preference for inclusive interpretations and then gradually develop a preference for exclusive interpretations as they grow older (Sternberg, 1979; Braine and Rumain, 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, giving the misleading impression that they are adopting inclusive interpretations. A similar phenomenon has been observed in several previous studies (Newstead et al., 1984; Roberge, 1977; Wason and Johnson-Laird, 1969) in which, under certain conditions, adult reasoners appeared to confuse the principles of "and" and "or" by using the terms synonymously.

Once the correct interpretation has been adopted, previous studies suggest that people generally find it easier to reason about exclusive disjunctives rather than inclusive disjunctives (Newstead et al., 1984; Newstead and Griggs, 1983; Roberge, 1977; 1978). Two explanations for this apparent difference in complexity are commonly offered. Firstly, it might be that people are generally more adept at reasoning with exclusive disjunctives simply because this is the more common form in everyday language. Secondly, it might be due to that 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 not sufficient for determining the truth value of the other. Nevertheless, it would appear that reasoning performance becomes significantly enhanced when it is clarified to reasoners which interpretation is to be drawn. This has typically been achieved in the past through the use of additional words in the disjunctive rules: "*p or q* (or both)" for inclusive disjunction, "*p or q* (but not both)" for exclusive disjunction.

Despite its frequent occurrence in everyday language, mathematical logic appears to have developed an almost universal bias against exclusive disjunction and tends to favour the inclusive form alone. The extent of this bias is typified in standard propositional logic, which contains an operator for expressing inclusive disjunction but no corresponding operator for exclusive disjunction. Although the propositional operator "$\vee$" was derived from the Latin term for inclusive disjunction, vel, it seems peculiar 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. Firstly, it is advantageous from a parsimonious perspective because all other logical operations can be defined in terms of inclusive disjunction and negation. Secondly, the inclusive operator neatly complements the set theoretic union operator because both refer to either one of two propositions or sets, and possibly both.[2]

[2]Nevertheless, logicians often invent compensatory non-standard symbols, such as Diller (1994) who defines the "$||$" symbol to denote exclusive disjunction.

$$
\begin{array}{lll}
p \text{ or } q & & \\
\text{not } p & & \\
\overline{\phantom{xxxx}} & Or \text{ denial} & \\
q & &
\end{array}
\qquad
\begin{array}{ll}
p \text{ or } q & \\
p & \\
\overline{\phantom{xxxx}} & Or \text{ affirmation} \\
\text{not } q &
\end{array}
$$

Figure 2: Denial and affirmation inferences for *Or*

The first argument shown in Figure 2 is often referred to as a "denial inference" because the minor premiss explicitly denies a term from the major premiss, resulting in the affirmation of the other term. The inference is logically valid under either an inclusive or an exclusive interpretation of the disjunctive rule in the major premiss. The second argument in Figure 2 is referred to as 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 rule and would be indeterminate under an inclusive interpretation (since both disjuncts might be true). Several natural language based studies conducted by Evans et al. (1993) suggest that the denial inference is made correctly around 84% and 80% of the time for exclusive and inclusive disjunctives, respectively. Their studies also suggest that around 83% of participants draw the affirmation inference correctly for exclusive disjunctives, but that around 36% persist in drawing it for inclusive disjunctives where it constitutes a logical fallacy. The experimenters argue 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, and that people have a general preference for determinate conclusions.

## Error and Bias in Conjunctive Reasoning

In contrast with conditionals and disjunctives, there has been relatively little cognitive research aimed at investigating the ways in which people reason about conjunctives. This might be due to the possibility that the linguistic rules and conventions governing the use of "and" are relatively simple and that people may be less prone to error and bias during conjunctive reasoning. However, Lakoff (1971) argues 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 terms, and this may be either overtly present or derivable by presupposition and deduction. He points to the existence of a hierarchy of sentences conjoined by "and" with varying strengths of relation between their common topics. At the top of this hierarchy are sentences like "John eats apples and he eats pears", where the meaning of one conjoined sentence complements the meaning of the other. At the bottom are 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. Lakoff, for example, argues 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. Those systems of formal logic which are defined in terms of the principles underlying Gentzen's (1935/1969) logical deductive calculus include two kinds of inference rule for connecting formal chains of reasoning: those for introducing and those for eliminating propositional connectives. Figure 3 shows the introduction and elimination rules for conjunction.

$$\frac{p \quad q}{p \text{ and } q}\ \textit{And}\text{ intro} \qquad \frac{p \textit{ and } q}{p}\ \textit{And}\text{ elim 1} \qquad \frac{p \textit{ and } q}{q}\ \textit{And}\text{ elim 2}$$

Figure 3: Logical rules of inference for *And*

Probability theory states that the likelihood of a conjunction, *p and q*, cannot exceed the likelihood of one of its constituent outcomes, $p$ or $q$. A series of studies conducted by Tversky and Kahneman (1983) sought to uncover conflicts between logic and intuition when reasoning about conjunctions. Specifically, the experimenters aimed to test whether people's systematic violations of this principle persisted across a variety of different contexts. Their results suggest that violation is likely whenever reasoners depart from principles of logic and begin adhering to intuitive heuristics, such as those of representativity and availability. The representativeness heuristic states that people are likely to judge an 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 an individual category. Of course, the main empirical issue to be addressed is whether the kinds of intuitive heuristics observed by Tversky and Kahneman are also adopted by trained logicians and, hence, whether the conjunctive fallacy transfers into the domain of formal reasoning.

$$\frac{\textit{not (p or q)}}{\textit{not p and not q}}\ \textit{Not}\text{ over }\textit{Or} \qquad \frac{\textit{not (p and q)}}{\textit{not p or not q}}\ \textit{Not}\text{ over }\textit{And}$$

Figure 4: De Morgan's laws

Figure 4 shows de Morgan's laws (Diller, 1994; Lemmon, 1993) which are used in formal reasoning to convert disjunctions into conjunctions and vice versa. Although they are perhaps not as commonly used as propositional logic's more basic rules, such as the introduction and elimination rules for propositional connectives, most trained logicians will almost certainly have used them beforehand. Of course, the question of whether they see their relevance in a given problem situation is another matter. The first argument shows how de Morgan's law can be applied to a negated disjunction to derive two separately negated conjuncts. The second argument shows how de Morgan's law can be applied to a negated conjunction to derive two separately negated disjuncts. It would be reasonable to expect that reasoners who have acquired a fair degree of deductive competence, especially those with prior training in mathematical logic, would have an appreciation of de Morgan's laws, and would be capable of combining them with other rules of inference in a chain of formal reasoning. For example, from the premisses "$\neg(p \wedge q)$" and "$p$", it is possible to deduce the logical conclusion "$\neg q$". This is achieved, firstly, by applying de Morgan's law over conjunction and then, secondly, by applying the rule for disjunctive elimination. In view of the relative complexity of this two-stage inference, it was an aim of the present study to test whether only experienced logicians are capable of mentally chaining together these kinds of inference in a chain of formal reasoning.

## The Mental Logic Theory of Reasoning

> "Deductive reasoning consists in the application of mental inference rules to the premises and conclusion of an argument. The sequence of applied rules forms a mental proof or derivation of the conclusion from the premises, where these implicit proofs are analogous to the explicit proofs of elementary logic."
>
> Rips (1983, p.40).

Recent cognitive research has propounded numerous theories which argue that people are fundamentally logical in nature and that reasoning is, to an extent, based on procedures akin to those found in systems of formal logic (Braine, 1978; Braine et al., 1984; Henle, 1962; Macnamara, 1986; Osherson, 1975; Rips, 1983). Proponents of the theory of "mental logic" assume that the human mind is equipped with a built-in set of abstract inference rules which are used for constructing mental derivations of conclusions in a wide range of problem contexts. It seems universally accepted amongst proponents of the theory that the types of inference rule stored in reasoners' mental repertories are analogous to the formal rules of inference found in formal systems of deductive logic, which includes rules for the propositional connectives and quantifiers. Mental logic 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 is true, then not only should the propositional inference rules for operator introduction and elimination be part of every person's mental machinery, but they should be two of the more basic rules at the formal logician's command. It was an aim of this research to test this hypothesis under a formal context for the disjunctive and conjunctive rules of inference.

> "It is tempting to attribute the difficulty of disjunctive concepts to the fact that the word 'or' is so ambiguous in the English language. ... 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, p.100).

It is hard to envisage a language that could be much more precisely defined than one whose syntax and semantics are defined entirely in terms of mathematical rules. Yet this is exactly the way in which formal notations are defined. In formal logic, disjunctives and conjunctives are not constrained by the same linguistic rules and conventions that govern their use in the English language. Formal expressions, for example, 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, such as drawing an exclusive interpretation when an inclusive interpretation is called for, and vice versa. However, the extent to which previous cognitive studies suggest that people are inherently biased towards exclusive interpretations in everyday language gives cause to suggest that people might still reason according to exclusive principles when they encounter inclusive disjunctives in formal logic. If this is the case in reality, then are formal logicians being asked to distort their natural thought processes and abandon intuitive judgement in favour of reason according to formally defined rules alone? These are the kinds of issue that the present study was aimed at illuminating.

## EXPERIMENT

### Method

*Participants.* Forty computer scientists and computing professionals from various academic institutions and industrial organisations volunteered to participate in the experiment. These were divided into two linguistic groups: Abstract Formal Logic (AFL) and Thematic Formal Logic (TFL). The groups were counter balanced, firstly, according to participants' personal ratings of Z expertise and, secondly, according to their lengths of Z experience. The AFL group comprised 14 staff, 3 students and 3 software professionals. Their mean age was 34.50 years ($s = 10.53$) and all had studied at least one system of mathematical logic beforehand (for example, the propositional or predicate calculus, Boolean algebra or Higher Order Logic). Their mean level of Z experience was 5.69 years ($s = 4.47$). According to participants' personal ratings of expertise, the group comprised 4 novice, 9 proficient and 7 expert users of the Z notation. The TFL group comprised 11 staff, 4 students and 5 software professionals. Their mean age was 35.00 years ($s = 9.85$) and 19 had studied at least one system of mathematical logic beforehand. Their mean level of Z experience was 4.67 years ($s = 3.80$), and the group comprised 4 novice, 9 proficient and 7 expert users.

*Design.* The study had a four factor design. The first, between groups, factor was the degree of realistic material: AFL and TFL. The second, repeated measures, factor was the type of inference to be drawn and had six levels: disjunctive elimination with a denied component (DE-D), disjunctive elimination with an affirmed component (DE-A), conjunctive elimination (CE), conjunctive introduction (CI), de Morgan's over disjunction with conjunctive elimination (DMD-CE), and de Morgan's over conjunction with disjunctive elimination (DMC-DE). The third, repeated measures, factor was the polarity of the logical premisses shown and had four levels for each of the DE-D, DE-A and CE inferences: AA, AN, NA and NN (where A and N correspond to the position of affirmative and negative components in the disjunctive or conjunctive premisses respectively). This factor also had two levels for the CI inferences: A and N. The fourth, repeated measures factor, was the position of the logical term eliminated, introduced, denied or affirmed, and had two levels: first or second. Owing to its popularity amongst the software engineering community and its mathematical foundations in "standard" logic (i.e. propositional logic with first-order predicate extensions), the Z notation (Spivey, 1992) was used to express the experimental tasks. Two logical equivalent sets of twenty eight tasks containing abstract and thematic content were presented to two corresponding linguistic groups. All reasoning tasks were syllogistic in nature and based on disjunctive and conjunctive logic.

TABLE 1
Logical forms of the disjunctive denial and affirmation tasks

| Polarity | Term | DE-D | DE-A |
|---|---|---|---|
| AA | 1 | $p \vee q,\ \neg p\ \therefore q$ | $p \vee q,\ p\ \therefore \neg q$ |
| AA | 2 | $p \vee q,\ \neg q\ \therefore p$ | $p \vee q,\ q\ \therefore \neg p$ |
| AN | 1 | $p \vee \neg q,\ \neg p\ \therefore \neg q$ | $p \vee \neg q,\ p\ \therefore q$ |
| AN | 2 | $p \vee \neg q,\ q\ \therefore p$ | $p \vee \neg q,\ \neg q\ \therefore \neg p$ |
| NA | 1 | $\neg p \vee q,\ p\ \therefore q$ | $\neg p \vee q,\ \neg p\ \therefore \neg q$ |
| NA | 2 | $\neg p \vee q,\ \neg q\ \therefore \neg p$ | $\neg p \vee q,\ q\ \therefore p$ |
| NN | 1 | $\neg p \vee \neg q,\ p\ \therefore \neg q$ | $\neg p \vee \neg q,\ \neg p\ \therefore q$ |
| NN | 2 | $\neg p \vee \neg q,\ q\ \therefore \neg p$ | $\neg p \vee \neg q,\ \neg q\ \therefore p$ |

*Note:* The following abbreviation refers to the disjunctive elimination inferences: *<Denial or affirmation>-<Major premiss polarity>-<Term denied or affirmed>*.

The disjunctive elimination (DE) reasoning tasks involved either the denial (D) of a component or the affirmation (A) of a component from the major premiss by the minor premiss. The polarity of components in the major premiss and the position of the component affirmed or denied in the minor premiss were systematically varied. Table 1 illustrates the underlying forms of the sixteen disjunctive denial and affirmation tasks. It should be noted that, whilst the DE-D inferences are logically sanctionable, all of the DE-A inferences shown are fallacious.

TABLE 2
Logical forms of the conjunctive elimination and introduction tasks

| Polarity | Term Order | CE | Polarity | Term Order | CI |
|---|---|---|---|---|---|
| AA | 2 | $p \wedge q \ \therefore p$ | A | 1 | $p \ \therefore p \wedge q$ |
| AN | 1 | $p \wedge \neg q \ \therefore \neg q$ | A | 2 | $q \ \therefore p \wedge q$ |
| NA | 2 | $\neg p \wedge q \ \therefore \neg p$ | N | 1 | $\neg p \ \therefore \neg p \wedge q$ |
| NN | 1 | $\neg p \wedge \neg q \ \therefore \neg q$ | N | 2 | $\neg q \ \therefore p \wedge \neg q$ |

*Note:* The following abbreviation refers to the conjunctive elimination inferences: CE-<*Premiss polarity*>-<*Term eliminated*>. The following abbreviation refers to the conjunctive introduction inferences: CI-<*Premiss polarity*>-<*Position of major term in conclusion*>.

The conjunctive reasoning tasks involved either the elimination (CE) or the introduction (CI) of logical components. The polarity of components in the arguments' premisses and the order of component introduced or eliminated were systematically varied. Table 2 shows the logical forms of the eight conjunctive elimination and introduction inference tasks. It should be noted that, whilst the CE inferences are logically sanctionable, all of the CI inferences shown are fallacious.

TABLE 3
Logical forms of the tasks requiring de Morgan's law

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

*Note:* The following abbreviation refers to the de Morgan's inferences: <*De Morgan's inference type*>-<*Disjunctive or conjunctive elimination*>-<*Term eliminated*>.

In addition to those reasoning tasks requiring the application of disjunctive or conjunctive rules in isolation, participants were presented with tasks which required the application of two rules of inference: de Morgan's over disjunction followed by conjunction elimination (DMD-CE), and de Morgan's over conjunction followed by disjunctive elimination (DMC-DE). Owing to the additional complexity of these inferences, the polarity of their premiss terms were held constant and only the type of component eliminated was varied. The underlying structures of these tasks are illustrated in Table 3.

*Task and Materials.* For each task participants were shown: the formal specification of a different software operation in the form of a Z operational schema, four statements in the form of Z predicates labelled "(*A*)" to "(*D*)" and a prompt describing information relevant to the operation. Participants were asked to select the one statement that followed from the task information by circling the letter corresponding to their response. In order to minimise the possibility of interference from meaningful content in the abstract group, a colours and shapes scenario was employed throughout. The abstract group was also told that they may assume the existence of the Z type definitions shown in Appendix A, which merely clarified the range of values that the operations' variables could be assigned. In contrast, the thematic group were presented with formal specifications of imaginary but realistic systems. These scenarios included a missile guidance system, a nuclear reactor cooling system, a live event's television coverage, a telephone network and a hotel reservation system. In order to minimise any potential conflict between logic and prior belief, all tasks were designed to lead to believable conclusions; that is, to plausible conceptions of the corresponding real-world applications. However, in order to avoid the correct answers simply being "read off" from memory with no employment of reasoning processes, more than one plausible conclusion was included in the available response options. The experiment's materials are exemplified in Appendix B of this paper. Following each task, participants were asked to give a subjective rating of the extent to which they believed their response was correct. This was achieved by ticking an appropriate box, as shown below. All task sheets were computer generated.

Confidence rating: ☐ Not confident ☐ Guess ☐ Confident

*Procedure.* Prior to the experiment, participants were 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. Both groups were then shown the instructions shown below.

> "In each of the tasks that follow, you will be shown a Z operational schema and a description of the operation's execution. You will be asked to determine which one of four given statements follow from the information given. Please circle the letter of your choice. You will also be asked to give a confidence rating, which should indicate how far you believe your answer to be correct. Please complete all tasks to the best of your ability, without reference to textbooks. The experiment should take no longer than 30 minutes to complete."

Task sheets were distributed to participants and completed anonymously, then mailed back to the experimenter. All participants were tested individually.

## Results

TABLE 4
Frequencies of valid and fallacious inferences endorsed in the two groups

| Group AFL ($n$ = 20) | | | | Group TFL ($n$ = 20) | | | |
|---|---|---|---|---|---|---|---|
| DE-D-AA-1 | 19 | DE-A-AA-1 | 20 (0) | DE-D-AA-1 | 19 | DE-A-AA-1 | 16 (4) |
| DE-D-AA-2 | 20 | DE-A-AA-2 | 18 (2) | DE-D-AA-2 | 19 | DE-A-AA-2 | 18 (2) |
| DE-D-AN-1 | 18 | DE-A-AN-1 | 18 (2) | DE-D-AN-1 | 18 | DE-A-AN-1 | 18 (1) |
| DE-D-AN-2 | 19 | DE-A-AN-2 | 19 (0) | DE-D-AN-2 | 19 | DE-A-AN-2 | 16 (4) |
| DE-D-NA-1 | 20 | DE-A-NA-1 | 17 (1) | DE-D-NA-1 | 18 | DE-A-NA-1 | 16 (1) |
| DE-D-NA-2 | 17 | DE-A-NA-2 | 18 (2) | DE-D-NA-2 | 18 | DE-A-NA-2 | 18 (0) |
| DE-D-NN-1 | 19 | DE-A-NN-1 | 18 (1) | DE-D-NN-1 | 17 | DE-A-NN-1 | 18 (2) |
| DE-D-NN-2 | 19 | DE-A-NN-2 | 18 (2) | DE-D-NN-2 | 18 | DE-A-NN-2 | 18 (2) |
| CE-AA-2 | 20 | CI-A-1 | 20 (0) | CE-AA-2 | 19 | CI-A-1 | 18 (2) |
| CE-AN-1 | 19 | CI-A-2 | 19 (1) | CE-AN-1 | 19 | CI-A-2 | 16 (2) |
| CE-NA-2 | 19 | CI-N-1 | 20 (0) | CE-NA-2 | 16 | CI-N-1 | 19 (1) |
| CE-NN-1 | 20 | CI-N-2 | 19 (0) | CE-NN-1 | 19 | CI-N-2 | 17 (0) |
| DMD-CE-1 | 19 | DMC-DE-1 | 19 | DMD-CE-1 | 16 | DMC-DE-1 | 14 |
| DMD-CE-2 | 20 | DMC-DE-2 | 19 | DMD-CE-2 | 18 | DMC-DE-2 | 14 |

*Note:* Numbers in parentheses denote the frequencies of fallacious inferences.

Table 4 shows the frequencies of valid and fallacious inferences endorsed by participants from the two linguistic groups. An analysis of variance revealed no significant main effect of group type on correctness with the disjunctive inferences, a main effect of group type on correctness with the conjunctive inferences approaching significance ($F_{(1,38)} = 3.85, p = 0.06$), and a significant main effect of group type on correctness with the de Morgan's based inferences ($F_{(1,38)} = 5.08, p = 0.03$). Comparison of the mean correctness scores for the two groups revealed that the AFL group ($\bar{x} = 26.55, s = 9.13$) outperformed the TFL group ($\bar{x} = 24.35, s = 8.66$) overall. This performance difference is illustrated in Table 5, which shows the derived probabilities at which members of each group appeared to draw valid and fallacious inferences.

TABLE 5
Probabilities of drawing correct and fallacious inferences ($0 \leq p \leq 1$)

| Inference Type | Group AFL | Group TFL |
| --- | --- | --- |
| Correct Conjunctive | 0.98 (0.16) | 0.89 (0.30) |
| Correct Disjunctive | 0.93 (0.25) | 0.88 (0.32) |
| Correct De Morgan's | 0.96 (0.19) | 0.78 (0.42) |
| Fallacious Conjunctive | 0.01 (0.11) | 0.06 (0.24) |
| Fallacious Disjunctive | 0.06 (0.23) | 0.10 (0.30) |

*Note:* Standard deviations are shown in parentheses.

An analysis of variance revealed no significant effects of inference type on participants' correctness for the disjunctive and conjunctive inferences, but a significant main effect of inference type on participants' abilities to draw the de Morgan's based inferences ($F_{(1)} = 9.04, p = 0.05$). A further analysis of variance suggested that there were no significant effects of inference type on participants' proneness to commit the DE-A or CI fallacies. Comparison of the frequencies of correctness for each inference type suggested an overall rank order of difficulty as follows: CE < CI < DE-D < DMD-CE < DE-A < DMC-DE.

An analysis of variance revealed no significant effects of term polarity on participants' correctness or their proneness to fallacies for either the disjunctive or conjunctive inferences. There were, however, significant interactions between polarity and inference type in participants' conjunctive reasoning performance ($F_{(3,114)} = 2.93, p = 0.04$), and between polarity and group type in participants' proneness to the disjunctive affirmation fallacies ($F_{(3,114)} = 3.17, p = 0.03$). Comparison of the rates of correctness for each polarity type revealed an rank order of difficulty as follows: AA < NN = NA < AN for the DE-D inferences, AA = NN = AN < NA for the DE-A inferences, AA = NA < NN < AN for the CE inferences, and A < N for the CI inferences.

An analysis of variance revealed that the order of logical terms denied, affirmed, introduced or eliminated had no significant effects on participants' reasoning performance. Although the differences between term orderings were not statistically significant for any of the inference types under scrutiny, comparison of the frequencies of correctness suggested overall rank orders of difficulty as follows: DE-D-2 < DE-D-1, DE-A-2 < DE-A-1, CE-1 < CE-2 and CI-2 < CI-1.

TABLE 6
Mean confidence ratings for the two groups

| Group AFL ($n$ = 20) | | | | Group TFL ($n$ = 20) | | | |
|---|---|---|---|---|---|---|---|
| DE-D-AA-1 | 3.00 | DE-A-AA-1 | 2.95 | DE-D-AA-1 | 2.75 | DE-A-AA-1 | 2.85 |
| DE-D-AA-2 | 3.00 | DE-A-AA-2 | 2.95 | DE-D-AA-2 | 2.85 | DE-A-AA-2 | 2.85 |
| DE-D-AN-1 | 3.00 | DE-A-AN-1 | 2.90 | DE-D-AN-1 | 2.90 | DE-A-AN-1 | 2.75 |
| DE-D-AN-2 | 2.90 | DE-A-AN-2 | 3.00 | DE-D-AN-2 | 2.85 | DE-A-AN-2 | 2.85 |
| DE-D-NA-1 | 3.00 | DE-A-NA-1 | 2.90 | DE-D-NA-1 | 2.85 | DE-A-NA-1 | 2.85 |
| DE-D-NA-2 | 3.00 | DE-A-NA-2 | 2.90 | DE-D-NA-2 | 2.90 | DE-A-NA-2 | 2.95 |
| DE-D-NN-1 | 3.00 | DE-A-NN-1 | 2.90 | DE-D-NN-1 | 2.85 | DE-A-NN-1 | 2.85 |
| DE-D-NN-2 | 3.00 | DE-A-NN-2 | 2.95 | DE-D-NN-2 | 2.85 | DE-A-NN-2 | 2.85 |
| CE-AA-2 | 3.00 | CI-A-1 | 2.90 | CE-AA-2 | 2.75 | CI-A-1 | 2.75 |
| CE-AN-1 | 2.95 | CI-A-2 | 2.90 | CE-AN-1 | 2.95 | CI-A-2 | 2.75 |
| CE-NA-2 | 3.00 | CI-N-1 | 3.00 | CE-NA-2 | 2.80 | CI-N-1 | 2.80 |
| CE-NN-1 | 2.90 | CI-N-2 | 3.00 | CE-NN-1 | 2.95 | CI-N-2 | 2.75 |
| DMD-CE-1 | 2.95 | DMC-DE-1 | 2.90 | DMD-CE-1 | 2.90 | DMC-DE-1 | 2.85 |
| DMD-CE-2 | 2.90 | DMC-DE-2 | 2.90 | DMD-CE-2 | 2.80 | DMC-DE-2 | 2.95 |

*Note:* All confidence ratings range from 1.00 (not confident) to 3.00 (confident).

Judging by the confidence ratings shown in Table 6, most participants were highly confident in the correctness of their responses for all task types. An analysis of variance failed to reveal any significant effects or interactions of the main variables on participants' confidence ratings for any of the types of inference under scrutiny. However, several correlations between confidence and correctness are evident. For example, nearly 90% of the AFL group's confidence ratings are higher than or equal to the ratings for the corresponding inferences in the TFL group, which correlates with the finding that the abstract group outperformed the thematic group overall. Unlike the relatively sporadic rates of the TFL group, the AFL group's confidence ratings are consistently high across all types of inference. Assuming a correlation between participants' confidence and correctness, this pattern in the observed results could be explained by the possibility that participants had employed a fixed interpretative strategy across all of the abstract tasks, yet contemplated different interpretations of the terms involved in the thematic tasks which led to more inconsistent responses. A similar trend was observed in a comparison of conditional reasoning performance in abstract and thematic material (Staudenmayer, 1975).

An analysis by linear regression revealed a significant correlation between participants' ratings of expertise and their levels of correctness ($R = 0.34$, $F_{(1,39)} = 5.08, p = 0.03$), a correlation between participants' years of Z experience and their levels of correctness approaching significance ($R = 0.29$, $F_{(1,39)} = 3.55, p = 0.07$), but no correlation between participants' ages and their levels of correctness. Taken together, these results suggest that it was participants' increased usage and familiarity with the Z notation, rather than their increased ages, that was responsible for their improved levels of performance.

## Discussion

> "The point is that the ordinary language connective *or* does not always correspond neatly to the inclusive and exclusive disjunctions of logic, and there are times when it seems to defy translation into any truth table pattern."
>
> Newstead et al. (1984).

Given a disjunction, "$p \vee q$", and the assertion of one of its disjuncts, "$p$", an exclusive interpretation allows us to deduce the falsity of the other, $\neg q$. However, under an inclusive interpretation, we cannot logically infer anything about the truth value of "$q$" because both disjuncts might be true. Although most participants responded correctly to the DE-A tasks, of the remainder who erred, nearly all gave responses which were consistent with exclusive interpretations of the disjunctive rules. In the light of Newstead and Griggs' (1983) hypothesis that disjunctive reasoning should be better in those languages where disjunctives are unambiguous, it is hard to imagine a disjunctive operator that could be more precisely defined as one whose grammar is specified purely in terms of formal logic. The finding that 6% of participants in the abstract group and 10% in the thematic group still appeared to adopt exclusive interpretations of the inclusive Z operator, in spite of its formally defined semantics, has disturbing implications for the practice of software development because it represents a likely source of erroneous reasoning. In comparison, Roberge (1976b; 1977; 1978) reports fallacious response rates for the affirmation task of around 20%, but significant performance improvements on those affirmation tasks where the disjunctives are explicitly marked as being inclusive or exclusive. In a later study (Evans et al., 1993), where an error rate of 36% was reported for the affirmation based inferences, the experimenters attribute participants' poor performance to a "propositional bias"; that is, a general preference for drawing determinate true or false, as opposed to indeterminate, conclusions.

Given a disjunction, $p \vee q$, and the denial of one of its disjuncts, $\neg p$, logic allows us to deduce the truth of the other disjunct, $q$. Inspection of the high rates of correctness for the DE-D inferences in both linguistic groups suggests that participants were generally able to draw this inference regardless of the polarity of terms and the position of the disjunct eliminated. However, the fact that 6% of responses were incorrect in the abstract group and 9% were incorrect in the thematic group suggests that participants still experienced some reasoning difficulties and that the presence of meaningful material led reasoners away from what was logically sanctionable. In comparison, previous natural language based studies involving the denial inference report significant effects of the degree of thematic material used (Newstead et al., 1984) and error rates ranging from around 20% (Evans et al., 1993; Roberge, 1976b; 1977; 1978) up to 87.5% (Johnson-Laird and Tridgell, 1972).

> "The conjunction error demonstrates with exceptional clarity the contrast between the extensional logic that underlies most formal conceptions of probability and the natural assessments that govern many judgements and beliefs."
>
> Tversky and Kahneman (1983, p.310).

According to the laws of probability, the likelihood of a proposition "$p$" cannot exceed the likelihood of a conjunction "*p and q*". Analysis of the responses to the CI inferences suggests that most participants succeeded in concluding that nothing follows in response to these tasks. The fact that a higher rate of participants were observed to commit the fallacy in the thematic group could be explained by the possibility that, in general, a conjunction of two realistic terms sharing a plausibly valid relation is more likely to be endorsed than a conjunction of two completely abstract terms sharing an arbitrary, unmeaningful relation. Notably, the tendency to commit this fallacy was strongest on those tasks in which a causal, rather than an arbitrary, relationship existed between the two conjuncts. For example, in "$current\_loc' = target\_loc? \wedge mission' = Success$" and "$applicant? \notin banned \wedge members' = members \cup \{applicant?\}$", the presupposition of the truth of the first conjunct appears to be necessary for an adequate understanding of the second. Both of these fallacious conclusions were endorsed by 10% of the thematic group. In contrast, in the conclusions "$print\_queue' = \langle\rangle \wedge \neg(printer\_status' = Online)$"

and "$\neg(\#register' > MaxStudents) \wedge \neg(student' \in register')$", the conjuncts do not appear to be as strongly interdependent and this is perhaps reflected in their lower rates of endorsement.

Given a conjunction, "*p and q*", application of the formal rule for conjunctive elimination allows us to conclude either one of the conjuncts, "$p$" or "$q$", in isolation. Judging by the high rates of correctness observed for the CE tasks, participants experienced little difficulty in drawing this inference, despite the variation of premiss polarity and order of the conjunct eliminated. Like several of the high response rates observed for the other forms of inference, this finding might also be attributed to the expression of the tasks in formal logic and participants' prior experience with Gentzen style deductive calculi. That the abstract group outperformed the thematic group slightly might be explained by the possibility that the realistic material elicited intuitive heuristics based on guesswork or association, similar to those employed in everyday reasoning, as opposed to logical lines of reasoning.

Based on the results of a study aimed at testing people's abilities to reason with positive and negative statements (Wason, 1959), Wales and Grieve (1969) argue that participants' failure to apply de Morgan's laws correctly resulted in many of their downfalls. During the study, participants were asked to draw inferences from natural language based statements of the form "There is not both $p$ and $q$". That most gave responses consistent with an interpretation of the form "There is not $p$ and not $q$", rather than the more logically accurate "There is not $p$ or not $q$", appears to have led to the 80% error rate observed during the study's first trial. Clearly, the majority of Wason's participants had wrongly assumed that the negative applied to both of the conjuncts rather than just the first. But, given that most had no prior training in formal logic, perhaps it was not entirely reasonable of the experimenter to take for granted participants' appreciation of the necessary de Morgan's law that would have enabled them to complete the task successfully.

In order to derive the correct responses for the DMD-CE and DMC-DE tasks presented in the present study, participants were required to apply an appropriate de Morgan's law followed by application of an appropriate elimination rule. Clearly, the logical structure of these inferences are more complex than the other forms under scrutiny in the present study. If, as developmental studies have shown (Inhelder and Piaget, 1958; Neimark and Slotnick, 1970; Paris, 1973; Piaget, 1972) deductive competence increases with age, then one might expect the more experienced reasoners who have developed proficiency in logical deduction to perform well on these tasks. Although a ceiling effect was present, several analyses by linear regression revealed significant correlations between participants' correctness with the de Morgan's based inferences alone and, firstly, their years of Z experience ($R = 0.33$, $F_{(1,39)} = 4.76, p = 0.04$) and, secondly, their levels of Z expertise ($R = 0.32$, $F_{(1,39)} = 4.47, p = 0.04$). These results lend support to the hypothesis that the likelihood of drawing de Morgan's based inferences correctly increases along with the experience and expertise of the reasoner.

## Signs of Classical Error and Bias

The possible facilitating effects of thematic problem content has been the source of much contention in the cognitive science community. The debate has been particularly prominent during the past three decades in the light of numerous studies of conditional reasoning which suggest that reasoning performance is facilitated by embedding problem content in realistic material (see for example: Bracewell and Hidi, 1974; Gilhooly and Falconer, 1974; Griggs and Cox, 1982; Johnson-Laird et al., 1972; Wason and Shapiro, 1971). It is, however, a well supported finding that conclusions conforming with prior beliefs are more likely to be endorsed than those running contrary to prior knowledge, and that these statements are often endorsed at the expense of logical necessity (Barston, 1986; Evans et al., 1983; Henle and Michael, 1956; Janis and Frick, 1956; Morgan and Morton, 1944; Wilkins, 1928). In a previous study of deductive reasoning with different forms of logical rule (Van Duyne, 1974), strong correlations between the degree of realistic material used and conditional reasoning performance were noted. However, no significant correlations between degree of realistic material and reasoning performance were noted for the logically equivalent tasks expressed in terms of disjunctions and conjunctions. Taken together, the results of a previous study of conditional reasoning (Vinter et al., 1997) and the present study of disjunctive and conjunctive reasoning suggest that a similar situation exists in the formal domain. That is, the expression of disjunctive and conjunctive rules in thematic material does not appear to facilitate reasoning performance in the same ways observed for conditionals. In the present study, despite the fact that all of the thematic inferences led to intuitively plausible conclusions, the finding that higher rates of correctness were observed for many of the logically equivalent abstract inferences suggests that meaningful material actually had a depreciatory effect on reasoning performance. This finding is replicated in an earlier study of reasoning with inclusive disjunctives in natural language (Roberge, 1977).

The theory of "matching bias" claims that reasoners are liable to select, or evaluate as relevant, only those conclusions which contain one or more of the terms mentioned explicitly in the given premisses (Evans, 1972b; 1983a; 1983b). At first impression, the disjunctive statements "I will travel by car or by train" and "I will not travel by car or by train" both appear to concern the same objects: "car" and "train". So, when a reasoner is presented with a response option that fails to contain one or both of these terms, the theory predicts that he or she will judge that option as irrelevant, regardless of its actual logical validity. Although strong evidence of matching bias has been discovered in previous studies of conditional reasoning (Evans, 1972b; 1983a; 1983b; Van Duyne, 1974), no signs of the bias have been found in previous studies of disjunctive or conjunctive reasoning (Evans and Newstead, 1980; Van Duyne, 1974). This trend has, in general, been attributed to the possibility that matching bias is a special kind of associational bias that is highly sensitive to the linguistic form of the logical rule (Evans and Newstead, 1980). More specifically, Pollard and Evans (1981) argue that a conditional, "if $p$ then $q$", suggests a positive relationship between the two terms involved because one expects "$q$" to occur with "$p$". Whereas, a disjunctive, "$p$ or $q$", suggests a negative relationship because one expects "$p$" to occur without "$q$", and vice versa. On this basis, the experimenters attribute the bias to their frequent reference to positively associated events in everyday language. However, this theory does not explain the lack of support for matching bias in previous studies of conjunctive reasoning where there appears to exist a strong, positive relationship between the two logical terms, "$p$ and $q$".

Many of the correct responses to the tasks in the present study are also consistent with the predictions of matching bias because they involve bringing out one or more of the terms explicitly mentioned in the given premisses. So, in order to determine whether participants succumbed to the bias, we must focus our analysis on those tasks where significant proportions of the responses contained terms which were only implied in the given premisses. This is the case for the de Morgan's based inferences only. Owing to their form, in order for matching bias to have occurred on these tasks, instead of the correct conclusion being given, one would have expected to see a significant proportion of responses containing the negation of the correct conclusion. However, inspection of the responses for these tasks revealed that, of those participants who responded incorrectly to these tasks, nearly all gave responses claiming that nothing followed from the given premisses. This finding suggests that few, if any, participants succumbed to the bias during the present study. However, in a previous study of conditional based reasoning in formal logic

(Vinter et al., 1997), limited signs of matching bias were evident where participants appeared to select responses containing one or more of the terms explicitly mentioned in the given premisses for several types of inference in the abstract condition. Analysed together, these findings suggest that a similar situation in formal logic as that observed in natural language, albeit to a lesser extent, where matching bias is strongest on conditional inferences but relatively non-existent for disjunctive and conjunctive based inferences. These findings are consistent with the claims of Evans and Newstead (1980) that matching bias is a highly sensitive associational bias which depends on specific linguistic properties of the logical rule.

Several studies suggest that the polarity of components in disjunctive rules can have a dramatic influence on the complexity of reasoning tasks (Evans and Newstead, 1980; Johnson-Laird and Tridgell, 1972; Roberge, 1974; 1976a; 1976b; 1978; Wason and Johnson-Laird, 1969; 1972). In particular, the theory of "negative conclusion bias" (Evans, 1972c, 1977; 1993) argues that people are generally more inclined to endorse an inference whose conclusion is negative rather than affirmative because the former maximises the reasoner's chances of making a statement which is unlikely to be disproved. Normally, affirmative conclusions have particular referents, whereas negative conclusions have multiple referents, so cautious reasoners are liable to favour statements that make non-specific negative predictions over specific affirmative predictions, which are more likely to be refuted. The results of a previous study of reasoning with exclusive disjunctives (Roberge, 1976a) were, in fact, consistent with the predictions of the theory insofar as significantly more inferences with negative conclusions were drawn correctly. Had participants succumbed to this bias in the present study we would expect to have seen significantly more negative conclusions drawn than affirmative ones for each of the main types of inference. That no such trend is evident in the observed results, where roughly equal numbers of affirmative and negative conclusions were drawn correctly, suggests that participants exhibited no signs of this bias. Similarly, the theory of "affirmative premiss bias" (Evans, 1993) argues that people are generally more inclined to endorse determinate conclusions from premisses that contain only affirmative components. Analysis of the results reveals that participants were equally likely to endorse determinate and indeterminate conclusions for each type of inference, which suggests that participants also avoided succumbing to this bias. The relative lack of support for these polarity biases suggest that they are predominantly conditional reasoning biases that are specific to the grammatical form of the implicative rule.

In everyday communication, people are notoriously prone not to interpret a doubly negated statement as an affirmative and thereby exhibit a "double negative bias". Given a negated description of an object, "not blue", it can sometimes be difficult for an individual to see how a further negation, "not not blue", could result in the object becoming any less blue than it already is. This frequent disinclination in normal discourse to convert a doubly negated proposition into an affirmative can perhaps account for a number of the errors made in strictly logical laboratory based studies. Roberge (1976b), in particular, reports that reasoners experience difficulties with disjunctive denial inferences where a negative disjunctive term in the major premiss is denied by an affirmative term in the minor premiss. Again, the fact that the rates of correctness for this form of denial inference in the present study were not significantly different to those for the other denial inferences suggests that participants did not experience the same kind of difficulties when reasoning with disjunctive statements in formal material. Even the low rates of correctness observed for the de Morgan's based tasks in the thematic group cannot be attributed to reasoning difficulties with negative components because high rates of correctness were observed for the logically equivalent tasks in the abstract group. Instead, it seems much more probable that the difference in performance rates observed for these inferences is attributable to the degree of realistic content used.

It is argued that people are generally more inclined to recognise that negatives deny prior positives, rather than the converse, because the function of negation in everyday language is to deny plausible misconceptions (Evans, 1972a; 1972b; 1983a; 1983b; Roberge, 1978). In a previous study (Johnson-Laird and Tridgell, 1972), reasoners found it easier to draw denial inferences where the minor premiss comprised an explicit denial of a term from the major premiss, as opposed to an implicit denial. Figure 5 illustrates two forms of explicit and implicit denial inference presented to participants during the study.

| | | | |
|---|---|---|---|
| *p or q* | | *not p or q* | |
| *not p* | | *p* | |
| ——— | Explicit denial | ——— | Implicit denial |
| *q* | | *not q* | |

Figure 5: Explicit and implicit denial inferences

In the explicit denial example, a negative is used to deny a statement that might plausibly be true. The experimenters attribute participants' difficulties with the implicit denial example to the fact that, in everyday communication, the disjunct "*p*" would not have been negated in the first place if there had been reason to suspect that it might be true. Had participants exhibited a similar tendency to draw the denial inference only from explicit denials of disjuncts in the present study, one would expect to see significantly more correct inferences for the four forms of explicit DE-D denial than the four forms of implicit DE-D denial. However, no such trend is evident in the observed results. Nevertheless, the highly variable nature of participants' rates of correctness in the thematic group might be partly attributed to the possibility that negative operators make more or less plausible denials, depending upon the exact realistic term used.

Previous studies suggest that the presence of negative components increases the complexity of logical tasks (Evans, 1972c; Johnson-Laird and Tridgell, 1972; Wason, 1959). Yet, results from earlier studies of disjunctive reasoning (Evans and Newstead, 1980; Roberge, 1976b; 1978) suggest that arguments with one negative component in the disjunctive premiss are more difficult than those containing two. As a possible explanation for this phenomenon, it is theorised that people implicitly drop the negatives when they are present in both disjuncts because they find it easier to reason with affirmatives only. Hence, people's response patterns suggest that inferences involving two negatives are actually simpler than those containing just one because the latter cannot easily be converted into an affirmative form. However, the conversion of major premisses in this manner does not always lead to logically valid conclusions. If, for example, participants had converted premisses containing two negatives for the denial or affirmation inferences in the present study, this would have led to erroneous conclusions. However, the fact that similar numbers of participants responded correctly to the disjunctive inferences containing two affirmatives as the corresponding inferences containing two negatives throughout suggests that participants did not succumb to this form of illicit conversion, but were equally adept at reasoning with premisses containing two terms of the same polarity. Nevertheless, inspection of the scores for the disjunctive inferences are consistent with the hypothesis that reasoners experienced more difficulty with inferences in which the major premiss contained a mixture of polarities. In fact, the rank orders of difficulty for the disjunctive inferences are directly comparable with those observed by Roberge (1976a): AA $<$ NN $<$ NA $=$ AN. Furthermore, when the rank orders for disjunctive inference polarities are analysed together with those for the conjunctive elimination and introduction inferences, it becomes overwhelmingly clear that participants found it easier to reason with affirmative components, although no statistically significant differences are evident in this respect.

## CONCLUSIONS

The basic argument of mental logic theory is that reasoners who have acquired a high level of deductive competence would, under ideal circumstances, always utilise the appropriate rule from their repertoires of inference rules to enable them to reach the correct conclusion (Inhelder and Piaget, 1958; Rips, 1994). Assuming that the observed rates of correctness are a fair reflection on participants' logical reasoning performance, the findings of the present study appear to offer some degree of support for the theory. Nevertheless, there were still occasions on which participants appeared susceptible to logical error which might be somewhat surprising given their logical backgrounds and the fact that they were free to complete the tasks at their leisure, without the pressures often associated with laboratory based experiments. As a possible explanation for the erroneous responses observed, especially in the thematic group, it could be argued that participants analysed the formal expressions only at a syntactic level and assumed an informal semantics, possibly cued by the realistic nature of the material, similar to that which they use for the corresponding propositional connectives in everyday natural language. This explanation is supported by recent revisions of mental logic theory which claim that mental logic closely co-exists with procedures for drawing non-logical inferences (Braine, 1994; O'Brien, 1993; 1995) such as those based on pragmatics (Cheng and Holyoak, 1985; Levinson, 1983) and intuitive heuristics (Kahneman et al., 1991).

In light of the finding that a majority of reasoners appeared to reason correctly about disjunctive and conjunctive statements embedded in formal specifications, one might be forgiven for thinking that the implications for the process of software development are entirely favourable. But, despite the fact that the observed error rates are generally lower than those observed in previous natural language based studies, the possibility that the same numbers would fail to reason correctly about formal specifications in real-life situations could have ominous consequences, given the business- and safety-critical nature of some of the software applications currently being developed with formal methods. The possibility that even one software developer should fail to reason correctly about a safety-critical specification, let alone the 30% observed for certain kinds of inference in the present study, should therefore be a genuine reason for empirical concern.

Having recognised that software engineering has always been a human driven activity, and is likely to remain so in the foreseeable future, the software community is gradually beginning to address the human aspects of formal specification. This is reflected in the growing numbers of research programmes aimed at improving the the ways in which specifications are communicated to their potential audiences (Bainbridge et al., 1991; Finney, 1996; Gravell, 1991; Macdonald, 1991). Ongoing research at the University of Hertfordshire is focussing on those properties of formal specifications that correlate with human understanding and reasoning performance. If those exact combinations of grammatical construct which impair the judgement of software developers can be identified and compensatory measures introduced, then this may help to reduce the potential for human error in the software engineering process. After all, if we know when and where errors are most likely to occur then erroneous development decisions can be avoided and the numbers of defects introduced into delivered systems minimised.

## REFERENCES

Bainbridge, J., Whitty, R.W. and Wordsworth, J.B. (1991). Obtaining structural metrics of Z specifications for systems development. In J.E. Nicholls (Ed.), *Z User Workshop, Oxford, 1990, Workshops in Computing*, 269-281, London: Springer-Verlag.

Bracewell, R.J. and Hidi, S.E. (1974). The solution of an inferential problem as a function of stimulus materials. *Quarterly Journal of Experimental Psychology, 26*, 480-488.

Barston, J.L. (1986). *An investigation into belief biases in reasoning.* Unpublished PhD thesis, University of Plymouth.

Bowen, J.P. (1988). Formal specification in Z as a design and documentation tool. *Second IEE/BCS Conference, Software Engineering 88*, 164-168.

Braine, M.D.S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review, 85*, 1-21.

Braine, M.D.S. (1994). Mental logic and how to discover it. In J. Macnamara and G.E. Reyes (Eds.), *The Logical Foundations of Cognition*, Oxford: Oxford University Press.

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 Rumain, B. (1981). Development of comprehension of "or": Evidence for a sequence of competencies. *Journal of Experimental Child Psychology, 31*, 46-70.

Braine, M.D.S., Reiser, B.J. and Rumain, B. (1984). Some empirical justification for a theory of natural propositional logic. In G.H. Bower (Ed.), The *Psychology of Learning and Motivation, 18*, New York: Academic Press.

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

Diller, A. (1994). *Z. An Introduction to Formal Methods.* Second Edition, Chichester: Wiley and Sons.

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. (1972c). Reasoning with negatives. *British Journal of Psychology, 63*, 213-219.

Evans, J.St.B.T. (1977). Linguistic factors in reasoning. *Quarterly Journal of Experimental Psychology, 29*, 297-306.

Evans, J.St.B.T. (1983a). Linguistic determinants of bias in conditional reasoning. *Quarterly Journal of Experimental Psychology, 35A*, 635-644.

Evans, J.St.B.T. (1983b). Selective processes in reasoning. In J.St.B.T. Evans (Ed.), *Thinking and Reasoning. Psychological Approaches*, 135-163, London: Routledge.

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

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: Erlbaum.

Fillenbaum, S. (1974). Or: Some uses. *Journal of Experimental Psychology*, 103, 913-921.

Finney, K. (1996). Mathematical notation in formal specification: Too difficult for the masses? *IEEE Transactions on Software Engineering, 22*, 2, 158-159.

Gravell, A.M. (1991). What is good formal specification? In J.E. Nicholls (Ed.), *Z User Workshop, Oxford 1990, Workshops in Computing*, 137-150, London: Springer-Verlag.

Gehani, N. (1986). Specifications: Formal and informal - A case study. In N. Gehani and A.D. McGettrick (Eds.), *Software Specification Techniques*, Wokingham: Addison-Wesley.

Gentzen, G. (1935/1969). Investigations into logical deduction. In M.E. Szabo (Ed.), *The Collected Papers of Gerhard Gentzen*, Amsterdam, North-Holland.

Gilhooly, K.J. and Falconer, W.A. (1974). Concrete and abstract terms and relations in testing a rule. *Quarterly Journal of Experimental Psychology, 26*, 355-359.

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.

Henle, M. (1962). On the relation between logic and thinking. *Psychological Review, 69*, 366-378.

Henle, M. and Michael, M. (1956). The influence of attitudes on syllogistic reasoning. *Journal of Social Psychology, 44*, 115-127.

Hurford, J.R. (1974). Exclusive or inclusive disjunction. *Foundations of Language*, 11, 409-411.

Ince, D. (1992). *An Introduction to Discrete Mathematics, Formal System Specification, and Z.* Oxford: Clarendon Press.

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.

Janis, L. and Frick, F. (1943). The relationship between attitudes toward conclusions and errors in judging logical validity of syllogisms. *Journal of Experimental Psychology, 33*, 73-77.

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

Kahneman, D., Slovic, P. and Tversky, A. (1991). *Judgement Under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press, 1991.

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.

Levinson, S.C. (1983). *Pragmatics.* Cambridge: Cambridge University Press.

Liskov, B. and Berzins, V. (1986). An appraisal of program specifications. In N. Gehani and A. McGettrick (Eds.), *Software Specification Techniques*, Wokingham: Addison-Wesley.

Loomes, M.J. and Vinter, R.J. (1997). Formal methods: No cure for faulty reasoning. In F. Redmill and T. Anderson (Eds.), *Safer Systems. Proceedings of the Fifth Safety-critical Systems Symposium, Brighton, 1997*, 67-78, London: Springer-Verlag.

Macdonald, R. (1991). *Z usage and abusage.* Internal Report No. 91003, RSRE, Ministry of Defence, Malvern, Worcestershire, UK.

Macnamara, J. (1986). *A Border Dispute. The Place of Logic in Psychology*, Cambridge, Massachusetts: 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.

Neimark, E.D. and Slotnick, N.S. (1970). Development of the understanding of logical connections. *Journal of Educational Psychology, 61*, 451-460, 1970.

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.

Newstead, S.E., Griggs, R.A. and Chrostowski, J.J. (1984). Reasoning with realistic disjunctives. *Quarterly Journal of Experimental Psychology, 36A*, 611-627.

O'Brien, D.P. (1993). Mental logic and human irrationality. We can put a man on the moon, so why can't we solve these logical-reasoning problems? In K.I. Manktelow and D.E. Over (Eds.), *Rationality: Psychological and Philosophical Perspectives*, 110-135, London: Routledge.

O'Brien, D.P. (1995). Finding logic in human reasoning requires looking in the right places. In S.E. Newstead and J.St.B.T. Evans (Eds.), *Perspectives on Thinking and Reasoning. Essays in Honour of Peter Wason*, 189-216, Hove, UK: Erlbaum.

Osherson, D. (1975). Logic and models of logical thinking. In R.J. Falmagne (Ed.), *Reasoning: Representation and Process in Children and Adults*, 81-91, Hillsdale, NJ: Erlbaum.

Paris, S.G. (1973). Comprehension of language connectives and propositional logical relationships. *Journal of Experimental Child Psychology, 16*, 278-291.

Pelletier, F.J. (1977). Or. *Theoretical Linguistics, 4*, 61-74.

Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. *Human Development, 15*, 1-12.

Pollard, P. and Evans, J.St.B.T. (1981). The effect of prior beliefs in reasoning: An associational interpretation. *British Journal of Psychology, 72*, 73-82.

Potter, B., Sinclair, J. and Till, D. (1991). *An Introduction to Formal Specification and Z.* Hemel Hempstead: Prentice-Hall.

Rips, L.J. (1983). Cognitive processes in propositional reasoning. *Psychological Review, 90*, 38-71.

Rips, L.J. (1994). The Psychology of Proof: Deductive Reasoning in Human Thinking. Cambridge, Massachusetts: MIT Press.

Roberge, J.J. (1974). Effects of negation on adults' comprehension of fallacious conditional and disjunctive arguments. *Journal of General Psychology, 91*, 287-293.

Roberge, J.J. (1976a). Reasoning with exclusive disjunctive arguments. *Quarterly Journal of Experimental Psychology, 28*, 419-427.

Roberge, J.J. (1976b). Effects of negation on adults' disjunctive reasoning abilities. *Journal of General Psychology, 24*, 87-91.

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.

Sommerville, I. (1992). *Software Engineering.* Fourth Edition, Wokingham: Addison-Wesley.

Spivey, J.M. (1992). *The Z Notation: A Reference Manual.* Second Edition. Hemel Hempstead: Prentice-Hall.

Staudenmayer, H. (1975). Understanding conditional reasoning with meaningful propositions. In R.J. Falmagne (Ed.), *Reasoning: Representation and Process in Children and Adults*, 55-79, Hillsdale, NJ: Erlbaum.

Sternberg, R.J. (1979). Developmental patterns in the encoding and comprehension of logical connectives. *Journal of Experimental Child Psychology, 28*, 469-498.

Turner, G.W. (1986). *Stylistics.* Harmondsworth: Penguin.

Tversky, A. and Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review, 90*, 4, 293-315.

Van Duyne, P.C. (1974). Realism and linguistic complexity in reasoning. *British Journal of Psychology, 65*, 59-67.

Vinter, R.J., Loomes, M.J. and Kornbrot, D.E. (1996). *Reasoning about formal software specifications: An initial investigation.* University of Hertfordshire, Division of Computer Science, Technical Report No. 249, March 1996.

Vinter, R.J., Loomes, M.J. and Kornbrot, D.E. (1997). *Conditional reasoning in language and logic: Transfer of non-logical heuristics?* University of Hertfordshire, Division of Computer Science, Technical Report No. 276, March 1997.

Wales, R.J. and Grieve, R. (1969). What is so difficult about negation? *Perception and Psychophysics, 6*, 327-331.

Wason, P.C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology, 11*, 92-107.

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.

Wilkins, M.C. (1928). The effect of changed material on ability to do formal syllogistic reasoning. *Journal of Social Psychology, 24*, 149-175.

# APPENDIX A

## Additional Instructions

This appendix contains the additional Z definitions shown to the AFL group.

$$\begin{array}{l} SHAPE ::= square \mid circle \mid triangle \mid rectangle \\ COLOUR ::= red \mid green \mid blue \mid white \end{array}$$

# APPENDIX B

## Examples of the Materials Used

This appendix contains examples of the abstract and thematic inference tasks presented to participants in the present study.

Group AFL, Inference DE-D-AA-1.

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

$$\begin{array}{|l} \hline \textit{GetShapeColour} \\ shape! : SHAPE \\ colour! : COLOUR \\ \hline colour! = white \vee \neg(shape! = rectangle) \\ \hline \end{array}$$

| | |
|---|---|
| (A) $shape! = square$ | (C) $\neg(shape! = rectangle)$ |
| (B) $shape! = circle$ | (D) Nothing |

Group TFL, Inference DE-A-NN-1.

If $\neg(processor! = Pentium)$ after its execution, what can you say about $display!$ in operation $ComputerHardware$?

$$\begin{array}{|l} \hline \textit{ComputerHardware} \\ processor! : Chip \\ display! : Screen \\ \hline \neg(processor! = Pentium) \vee \neg(display! = HighResolution) \\ \hline \end{array}$$

(A) $display! = LowResolution$ (C) $display! = HighResolution$
(B) $\neg(display! = HighResolution)$ (D) Nothing

Group TFL, Inference CE-NA-2

What can you say about the effect of operation *HireVideo* on its after-state variables?

$$\begin{array}{|l} \hline \textit{HireVideo} \\ \Delta\, VideoShop \\ \hline \neg(film' \in FilmsOnShelf) \wedge report' = OnLoan \\ \hline \end{array}$$

(A) $film' \in FilmsOnShelf \wedge report' = OnLoan$ (C) $\neg(film' \in FilmsOnShelf)$
(B) $\neg(report' = OnLoan)$ (D) Nothing

Group TFL, Inference CI-A-1

What can you say about the effect of operation *GuidedMissileCheck* on its after-state variables?

$$\begin{array}{|l} \hline \textit{GuidedMissileCheck} \\ \Delta Bearings \\ target\_loc? : COORDS \\ \hline current\_loc' = target\_loc? \\ \hline \end{array}$$

(A) $\neg(current\_loc' = target\_loc?) \wedge mission' = Failure$ (C) $\neg(current\_loc' = target\_loc?)$
(B) $current\_loc' = target\_loc? \wedge mission' = Success$ (D) Nothing

Group TFL, Inference DMD-CE-2

Based on its description below, what can you say about the output from operation *TimeAndDate*?

$$\begin{array}{|l} \hline \textit{TimeAndDate} \\ time! : TimeSystem \\ calendar! : CalendarSystem \\ \hline \neg(time! = GMT \vee calendar! = Julian) \\ \hline \end{array}$$

(A) $time! = GMT$ (C) $\neg(time! = WET)$
(B) $\neg(time! = GMT)$ (D) Nothing

Group TFL, Inference DMC-DE-2

If $line\_status' = Unconnected$ after the execution of operation *ConnectNewUser*, what can you say about $user'$?

$$\begin{array}{|l} \hline \textit{ConnectNewUser} \\ \Delta\, TelephoneNetwork \\ \hline \neg(user' \notin ConnectedUsers \wedge line\_status' = Unconnected) \\ \hline \end{array}$$

(A) $\neg(user' \notin ListedUsers)$ (C) $user' \notin ConnectedUsers$
(B) $\neg(user' \notin ConnectedUsers)$ (D) Nothing