

**DIVISION OF COMPUTER SCIENCE**

**Quantified Reasoning in Formal Specification: Transfer  
of Everyday Errors and Biases?**

**R J Vinter, M J Loomes and D E Kornbrot**

**Technical Report No. 299**

**July 1997**

# Quantified Reasoning in Formal Specification: Transfer of Everyday Errors and Biases?

Rick Vinter and Martin Loomes  
*Faculty of Engineering and Information Sciences*

Diana Kornbrot  
*Faculty of Health and Human Sciences*  
*University of Hertfordshire, Hatfield, UK*<sup>1</sup>

## Abstract

Cognitive studies of syllogistic reasoning appear to provide important pointers to the ways in which people reason with quantified statements. Previous natural language based studies of the syllogistic task suggest that novice reasoners are prone to systematic errors and biases. In this report, we discuss some of the cognitive explanations for these and report a study aimed at testing whether computing scientists with the relevant training in logical deduction and the Z formal notation are liable to succumb to the same non-logical tendencies when reasoning about categorical syllogisms expressed in Z. The results suggest that many of the errors and biases which people exhibit on a frequent basis when reasoning about quantified statements in everyday communication can transfer over into the formal domain. The implications of this finding are discussed in relation to the software engineering community where formal specifications are becoming increasingly used in the development of business and safety-critical systems.

## 1 Introduction

In everyday life, the success of our decisions often relies upon the accuracy of the reasoning processes which lead to them. Normally, we are able to reach decisions instantaneously and automatically, unaware of the cognitive processes upon which they are based (Evans et al., 1993; James, 1950), and we receive almost immediate feedback informing us whether or not our decisions are the correct ones. If it transpires that our decision has guided behaviour badly then, providing it is not already too late, we are free to change our mind and set an alternative chain of events in motion. In the context of software engineering, however, the consequences of making inaccurate decisions can be rather less forgiving. Typically, most key development decisions are made at an early stage in a software project and developers do not receive feedback on their accuracy until near its completion, by which time it may transpire that these initial decisions were incorrect and that additional work is required to rectify the mistakes incurred. The entire history of software engineering has shown us that the cost and effort required to rectify such work tends to increase along with the time taken to recognise that those early development decisions were not the right ones (Cohen, 1989; Sheppard and Ince, 1993). It therefore seems critical to the success of software projects that early development decisions in particular are carefully deliberated and based on sound reasoning wherever possible. But, besides incurring the delays and costs associated with correcting erroneous work, incorrect development decisions can also lead to the introduction of defects

---

<sup>1</sup>The authors would like to thank Richard Ralley and Ken Ryder for their technical assistance, Carol Britton, Claire Millward, Ben Potter and Jane Simpson for reviewing this report, and 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 should be sent to Rick Vinter, Faculty of Engineering and Information Sciences, Hatfield, Hertfordshire, AL10 9AB, UK.

in "finished" software systems. Given the increasing complexity and criticality of software (MacKenzie, 1992), this is a genuine reason for concern.

Historically, failure to interpret or reason correctly with software specifications has caused developers to make incorrect development decisions which, in turn, have led to the introduction of faults or anomalies in software systems (Fenton and Pfleeger, 1996; Potter et al., 1996). The software community argues that the problems stem from designers' near exclusive use of natural language based notations which are notoriously prone to imprecision (Gehani, 1986; Meyer, 1985) and verbosity (Barroca and McDermid, 1992; Norcliffe and Slater, 1991). This claim is supported universally by the linguistic literature which reaffirms that natural language is inherently vague and ambiguous (Empson, 1965; Turner, 1986). It is also supported by the cognitive literature, which suggests that natural language sentences containing certain forms of logical connective are prone to incite human reasoning errors and, consequently, to cause erroneous decisions: "if" (Braine and O'Brien, 1991), "and" (Lakoff, 1971), "or" (Newstead et al., 1984), "not" (Johnson-Laird and Tridgell, 1972), "some" and "all" (Erickson, 1978; Johnson-Laird, 1977).

However, the past two decades have seen some striking new developments in software technology, perhaps the most notable of which has been the advent of "formal methods"; languages and tools with precise mathematical underpinnings designed to overcome some of the intrinsic weaknesses of natural language based notations. The advocates of formal methods claim distinct advantages over natural language including increased precision and concision, and increased levels of insight and confidence gained through the use of a mathematical approach (Ince, 1992; Liskov and Berzins, 1986). However, such claims are often based on personal conviction, anecdotal evidence or isolated case studies from which results can be difficult to generalise (Fenton, 1996). Moreover, several of these claims rest on psychological assumptions which have yet to be proven. The fact that empirical science has yet to produce substantive evidence which might refute the formalists' claims appears, for many, to support the case for formalisation. But, supposing that formal specifications are indeed more precise than their informal counterparts and that audiences are more likely to interpret them correctly, it follows that other benefits could follow as a consequence. Perhaps the most notable of these possible benefits is that developers would be more likely to reason correctly about formal specifications, at both a formal and informal level (Thomas, 1995). This possibility is worthy of special attention because improved reasoning could lead to more accurate development decisions which, in turn, could lead to the introduction of fewer defects in software systems.

Aside from being overwhelmingly appealing, the possibility that a prominent source of software defect could be eliminated through the process of formalisation appears, for many, to be intuitively plausible. Although the possibility that users are more likely to reason correctly with formal methods is not a claim that is often made explicitly in their favour, it is nonetheless an implicit assumption which appears prominent amongst a large proportion of the software community. The justification for this assumption appears to stem from the argument that it is generally easier to manipulate and reason about problems expressed in mathematical logic than those expressed in natural language (Ince, 1992; Lemmon, 1993). Reichenbach (1966, p.3) argues "It is true that simple logical operations can be performed without the help of symbolic representation; but the structure of complicated relations cannot be seen without the aid of symbolism". Intuitively, there appears to be no reason why this argument should not generalise to specification languages whose grammatical foundations lie in these domains. If this assumption is correct then the software community stands to make tremendous gains from the adoption of formal methods because they could provide a long awaited key to the development of safer systems.

But despite obvious syntactic differences, most formal notations contain logical

operators with roughly equivalent meanings as those same natural language constructs which have been shown to incite incorrect decisions in previous cognitive studies:  $\Rightarrow$  (if),  $\wedge$  (and),  $\vee$  (or),  $\neg$  (not),  $\exists$  (some) and  $\forall$  (all). Of course, the main question that the software community must ask itself 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 fact that people currently err at all in the specification process is a reason for concern, but the possibility that they will continue to do so even after having adopted a formal approach is especially disconcerting given the business and safety-critical nature of the projects to which they are applied (Barroca and McDermid, 1992; Bowen and Stavridou, 1993). Furthermore, if software developers assume that the use of formal methods will promote error-free reasoning when, in fact, this belief is inaccurate, then it is likely to instill a false sense of security. The present study is aimed at testing whether trained computing scientists are liable to succumb to these same errors and biases when reasoning with statements containing two such formal operators: the existential ( $\exists$ ) and universal ( $\forall$ ) quantifiers. In order to help us achieve this aim, we borrow from cognitive science relevant theoretical knowledge and experimental methodology.

## 1.1 Principles of Syllogistic Reasoning

The doctrine of the “syllogism” was originally propounded by Aristotle in Ancient Greece to model the simplest form of deductive chain and provide a content-independent means for reducing and evaluating the validity of arguments (Adams, 1984). For nearly two thousand years, the authority of Aristotle remained unquestioned while it was universally accepted that all deductive reasoning was reducible to syllogistic form and the syllogism was regarded as an accurate model of people’s deductive thought processes. It was not until the nineteenth century that the validity of Aristotle’s theories were brought into question and their weaknesses exposed (for a critique see: Beth and Piaget, 1966; Strawson, 1966). The possibility that Aristotelian logic might, then, serve as a candidate model for accurately representing human deductive thought processes has been widely dismissed in the cognitive literature mainly because syllogistic reasoning rarely seems to occur in everyday argumentation (Newstead, 1989), or in mathematics which is almost entirely deductive (Russell, 1994). Indeed, to translate arguments or mathematical problems into syllogistic form would be artificial and might even confound matters by disguising the nature of the original problem. Nevertheless, the syllogistic task is of special interest to cognitive science because it encompasses several of the core cognitive processes that appear to pervade human reasoning generally. The interpretation of premisses, the integration and representation of terms, the hypothetical postulation and evaluation of speculative conclusions, and the generation of responses are processes which are not confined to the syllogistic task, but instead appear to pervade many facets of human reasoning (Evans et al., 1993; Dickstein, 1978b). Indeed, if psychology proves unable to account for the cognitive determinants of performance in the syllogistic task, then it is difficult to see how it will ever come to explain more complex cognitive functions (Johnson-Laird and Bara, 1984). Studies of syllogistic reasoning therefore provide important pointers to the cognitive processes involved in human reasoning generally and, in particular, the ways in which people reason with quantified statements.

A categorical syllogism is an argument consisting of three statements: a major premiss, a minor premiss and a conclusion. Each of these statements describe relations between the various “terms” of the argument. The major premiss describes the relation that holds between the predicate of the conclusion (P) and a middle

term (M). The minor premiss describes the relation that holds between the subject of the conclusion (S) and the middle term. Convention states that the major premiss must always precede the minor premiss. The aim of the syllogistic task is to use the two premisses as the basis for deducing a conclusion which describes a relation that exists between S and P, or, where the premisses cannot lead to such a deduction, to state that no determinate conclusion follows. Four types of quantifier may range over the assertions made in a syllogism: "All", "Some", "Some ... not" and "No". The quantifier which ranges over a syllogistic predicate reflects that predicate's "mood", conventionally abbreviated as shown in Figure 1.

Universal affirmative	All M are P	(A)
Universal negative	No M are P	(E)
Particular affirmative	Some M are P	(I)
Particular negative	Some M are not P	(O)

Figure 1: The four moods of syllogistic predicate

The ordering of terms in a syllogism's premisses is significant. As there are two possible orderings for each of the major and minor premisses, this gives rise to four possible arrangements, or "figures", as shown in Figure 2. Although the order in which terms are presented within the two premisses might also vary, the ordering of terms in the conclusion always proceeds from S to P.

Figure 1	Figure 2	Figure 3	Figure 4
M-P	P-M	M-P	P-M
S-M	S-M	M-S	M-S
—	—	—	—
S-P	S-P	S-P	S-P

Figure 2: The four figures of a syllogism

Figure 3 shows an example of a syllogism with the form AA1; one which Aristotle would consider to be a "perfect" syllogism, that is, one whose necessity can be seen by novice reasoners without logical expertise (Adams, 1984; Lukasiewicz, 1967), and "one that needs nothing other than the premisses to make the conclusion evident" (Aristotle, in Ross, 1949, p.287). The syllogism in Figure 4 is of the form EO1. Here, the conclusion drawn is fallacious because one cannot say for certain whether this relation follows necessarily from the information specified in the given premisses. In this case, no determinate conclusion logically follows.

All humans are mortal	No Greeks are immortal
All Greeks are humans	Some men are not Greeks
—	—
All Greeks are mortal	Some men are not immortal

Figure 3: A "perfect" syllogism

Figure 4: An invalid syllogism

There is general agreement in the cognitive community that syllogistic reasoning involves three main cognitive stages, all of which are prone to error and bias: premiss interpretation, premiss combination and response generation (Erickson, 1974; Evans et al., 1993). In order to draw valid conclusions, it is imperative that reasoners adhere to logical principles of deduction. However, the results of previous syllogistic studies suggest that reasoners are often prone to depart from such principles and that there are dominant causes for their erroneous responses. The cognitive literature has been keen to speculate numerous possible explanations for these trends.

## 1.2 Error and Bias in the Syllogistic Task

A series of studies conducted by Woodworth and Sells (1935) gave rise to one of the earliest psychological theories of human error in the syllogistic task. According to "atmosphere theory", each pair of syllogistic premisses creates a certain global impression, or "atmosphere", depending upon how they are quantified and qualified. The quantity of a premiss can be universal ("all") or particular ("some"), whereas the quality of a premiss can be affirmative ("are") or negative ("are not"). The theory states that, whenever a reasoner is unable to see the link between S and P that will enable them to draw a valid conclusion, he or she will respond according to the atmosphere of the premisses. Begg and Denny's (1969) reformulation of atmosphere theory makes two specific predictions. Firstly, whenever the quality of at least one premiss is negative, the quality of the conclusion drawn will be negative; when both premisses are affirmative, the quality of the conclusion drawn will be affirmative. Secondly, whenever the quantity of at least one premiss is particular, the quantity of the conclusion drawn will be particular; when both premisses are universal, the conclusion drawn will be universal. In short, contemporary atmosphere theory predicts that, where the relationship between S and P is less than obvious, the reasoner will draw a conclusion which shares the same qualifiers and quantifiers as those contained in the premisses, with little or no regard for the logic of the syllogism.

The atmosphere effect, it appears, is not restricted to syllogistic reasoning. Sells (1936) claims that the atmosphere effect is influential in everyday problem solving, particularly where the range of possible solutions are limited. In such cases, the effect is claimed to lead the individual to endorse the solution "most similar to the general trend or tone of the situation set up" (p.7). A closely related theory is that of "matching bias" (Evans, 1972), which was originally used as an explanation for erroneous trends in studies of conditional reasoning but was later adapted to account for errors in the syllogistic task (Wetherick and Gilhooly, 1989; 1990). Like atmosphere theory, matching theory argues that reasoners are keen to avoid the logic of the syllogistic task. But, unlike atmosphere theory, it does not attempt to account for reasoning behaviour in response to all possible premiss combinations. Rather, it simply states that, where the reasoner is unsure of how to reach a conclusion via logical deduction, he or she will simply choose a conclusion whose quantitative form matches one of the two premisses. Normally, reasoners choose to match the more conservative premiss; that is, the one which makes an assertion about the fewest class members because, owing to the laws of probability, this form is least likely to be disproved. The results of a separate study (Johnson-Laird and Bara, 1984) suggest that reasoners are especially prone to matching bias when both premisses share the same quantitative form: AA, II, OO or EE.

Implicit conversion is an essential and regular feature of human comprehension in which people attempt to simplify complex problem information to forms that are more amenable to mental representation or processing (Revlín and Leirer, 1980). In the more specialised context of syllogistic reasoning, the theory of "implicit premiss conversion" argues that reasoners attempt to convert one or both premisses in a similar fashion. For example, the construction of transitive relations between a conclusion's end terms, S and P, can help to clarify the form of conclusion to be drawn. Illicitly converted forms, however, can form a false basis from which erroneous conclusions are drawn. Although conversion which does not fully adhere to the dictates of logic is often sufficient for the purposes of everyday reasoning, departure from logic in the context of the syllogistic task often leads to erroneous conclusions, unless of course the converted premisses lead fortuitously to the same logical conclusion as their original forms.

Logically, premiss conversion is permissible for the I and E forms because "Some S are M" can be freely replaced by "Some M are S", and "No S are M" can be freely replaced by "No M are S". However, conversion of the A and O forms is not logically permissible and frequently leads to error in the syllogistic task; "All S are M" does not necessarily imply "All M are S", and "Some S are not M" does not necessarily imply "Some M are not S". Previous syllogistic studies suggest that reasoners often fail to recognise the conditions under which conversion is acceptable and illicitly accept the converse of universal affirmatives or particular negatives at the expense of logical necessity (Chapman and Chapman, 1959; Dickstein, 1981; Newstead and Griggs, 1983; Wilkins, 1928). As part of a theoretical model of syllogistic reasoning, Revlis (1975a; 1975b; Revlin and Leirer, 1980) argues not only that reasoners have a tendency to convert premisses, but that conversion is actually the preferred method of interpreting syllogistic premisses. If this were true then reasoners would never attempt to reason from the same premisses that are presented to them! Strong evidence exists to suggest that this part of Revlis' model does not reflect the way in which most reasoners actually approach the syllogistic task (Johnson-Laird and Bara, 1984; Newstead and Griggs, 1983; Traub, 1977). Nevertheless, illicit premiss conversion would appear to account for a substantial proportion of the errors observed in syllogistic studies.

A brief review of the literature suggests that implicit premiss conversion can take several subtly different forms. Dickstein (1981) observes two variants of conversion theory. Conversion by addition occurs when, in addition to a given A or O premiss, a reasoner assumes its converse; for example, given "All A are B", a reasoner is liable to assume an interpretation of the form "All A are B, and all B are A." Conversion by substitution occurs when the given premiss is simply replaced by a converted one; for example, "All A are B" might become "Some B are A". Politzer (1990) recognises three additional types of premiss manipulation. Firstly, basic conversion occurs when subject and predicate simply exchange positions; for example, "All A are B" might become "All B are A". Secondly, contraposition occurs where subject and predicate exchange position and are negated or complemented; for example, "All A are B" might become "All non-B are non-A". Thirdly, obversion occurs where the predicate is negated or complemented and the quantity of the statement changes; for example, "No A are B" might become "All A are non-B."

A study conducted by Tsal (1977) suggests that reasoners tend to make unwarranted assumptions about the existence of symmetrical or transitive relations between terms connected by non-specified relations, especially where abstract problem material is used. That reasoners seem so strongly inclined to assume the existence of such relations should be worthy of our concern because it makes them particularly susceptible to conversion errors in the syllogistic task, where only two cases of true symmetry exist, "No A are B" implies "No B are A" and "Some A are B" implies "Some B are A", and two cases of limited symmetry exist, "All A are B" implies "Some A are B", and "No A are B" implies "Some B are not A". Although Tsal's study was aimed at testing human reasoning processes in general, his results are consistent with those of syllogistic studies. Chapman and Chapman (1959), for instance, suggest that the qualifier "are" encourages reasoners to assume the identity relation "is equal to" when the inclusion relation "is included in" would be more appropriate from a set-theoretic perspective. It is argued that this encourages reasoners to make unwarranted assumptions of symmetry, in a manner similar to that predicted by Tsal. The experimenters ascribe assumptions of this form to reasoners' prior experience of elementary mathematical algebra or geometry, where identity relations are commonplace.

Notably, Tsal's results suggest that the tendency to assume symmetrical or transitive relations disappears when a reasoning task is expressed in thematic content. This is supported by syllogistic studies which suggest that the use of thematic

content can sometimes block illicit conversion (Ceraso and Provitera, 1971; Chapman and Chapman, 1959; Evans et al., 1983; Henle and Michael, 1956; Newstead, 1990; Revlis, 1975a; Revlin et al., 1980). These findings might be explained by the possibility that, when premisses are couched in abstract material, a reasoner is unlikely to have strong dispositions towards the terms involved and is thus liable to regard them as being interchangeable with alternative forms. However, when the terms of the syllogism are couched in more meaningful content, conversion can lead to forms which conflict with a reasoner's conceptions of the real world. This might explain why, for example, reasoners seem less inclined to convert statements such as "All dogs are animals" into "All animals are dogs" (Evans et al., 1983; Henle and Michael, 1956; Newstead, 1990; Revlin et al., 1980).

Our knowledge of pragmatic laws and conventions guides our interpretation of written and spoken language, enabling us to look beyond what is said explicitly in order to gain an appreciation of a speaker's intentions (Moxey and Sanford, 1993). Although our pragmatic knowledge typically develops alongside our language acquisition skills from early adolescence onwards, our knowledge of logical principles does not fully develop until late adolescence or early adulthood (Inhelder and Piaget, 1958). Since the laws of logic and the laws of everyday communication are often incompatible, this development leads towards two contradictory systems and it is argued that the ability of reasoners to differentiate between them is a major determinant of reasoners' performance in deductive tasks (Politzer, 1986). The results of contemporary studies of syllogistic reasoning are often discussed specifically in relation to Grice's (1975) seminal work on conversational implicature. Under Grice's Cooperative Principle, which aims to explicate some of the universally accepted rules and conventions that govern everyday spoken and written communication, there are four maxims. Firstly, the maxim of quantity states that speakers should make their contribution as informative as is required for the purposes of the exchange. Secondly, the maxim of quality states that speakers should only say that which they believe to be true and supported by adequate evidence. Thirdly, the maxim of relation states that the speaker should keep the content of their contributions as relevant as possible. Fourthly, the maxim of manner states that the contributions made by speakers should be clear and unambiguous.

According to Grice, speakers formulate their verbal contributions around the four maxims on such a regular basis in everyday conversation that they have become part of the universally accepted norm. Furthermore, it is argued that listeners take for granted speakers' intentions to abide by the four maxims in order to help their interpretation of what is being said. Problems seem to arise, however, when people try to apply Gricean principles to reasoning tasks because the obligation to conform with these four maxims can lead reasoners away from the logical principles that underly the true nature of the task. After all, the pragmatic rules and conventions that govern everyday discourse in natural language are very different to the rules and conventions which tend to govern purely laboratory based studies of deductive reasoning. There is a wealth of evidence to suggest that reasoners are predisposed to apply the same Gricean conventions from everyday discourse to the syllogistic task (Begg and Harris, 1982; Newstead, 1989; 1995; Politzer, 1986).

Misapplication of Gricean maxims may be responsible for a large proportion of the erroneous responses observed in previous syllogistic studies. Whenever the subject and predicate of a speculative conclusion are related by a common middle term (i.e. the "Same M"), reasoners will tend to accept this conclusion at face value according to the maxim of relation, irrespective of its logical necessity. If in everyday conversation one were to say "Some politicians are lazy, and some lazy people are wealthy", then he or she is clearly inviting the listener to draw the conclusion "Some politicians are wealthy." Otherwise, the speaker could be accused of violating the maxim of relation and being deliberately deceitful. Although this



form of probabilistic inference frequently leads to correct conclusions in everyday experience, it does not follow from a logical perspective because the middle term in each premiss may not necessarily refer to the same class members. Strong evidence of this "Same M" fallacy has been reported in numerous syllogistic studies (Chapman and Chapman, 1959; Dickstein, 1975; 1976; Haviland and Clark, 1975). A similar phenomenon is documented by Dickstein (1978b), who argues that, in premisses where no relations are specified between S and M, and between P and M, a reasoner might still draw a conclusion from S to P because both premisses share the common property of not being related to M. The maxim of relation can therefore explain the tendency to give determinate conclusions in the syllogistic task where none are warranted because it is assumed that experimenters would not intentionally make two completely unrelated statements in sequence.

Reasoners' adherence to the Gricean maxim of quantity may explain a further source of error in syllogistic reasoning. It recommends that speakers make their contributions as informative as possible and that they do not withhold information they know to be true. Thus, a speaker who says "some" when they know "all" to be the case, or "some are not" when they know "none" to be the case, could be accused of violating this maxim. However, the syllogistic task often requires reasoners to entertain counter-intuitive notions, such as "Some apples are fruits", even when they know that in fact "All apples are fruits." Further violations of this maxim are evident in the results of Woodworth and Sells' (1935) study, where probabilistic inferences appeared to lead reasoners to adopt a non-logical "caution bias"; that is, a tendency to accept "Some ... are" conclusions more readily than "All ... are" conclusions, and "Some ... are not" conclusions more readily than "None ... are" conclusions. This frequent inclination to accept weak conclusions, when it is known that a stronger version might exist, suggests that reasoners regularly fail to consider all hypothetical possibilities and that they are generally conservative estimators.

The interpretational problems caused by discrepancies between the meaning of the quantifier "some" in ordinary language and its meaning in formal logic are well documented (Erickson, 1974; Wilkins, 1928; Woodworth and Sells, 1935). Formal logic defines "some" to mean "at least one, and possibly all" and syllogistic errors are often attributed to the possibility that reasoners confuse this meaning with its more common meaning in everyday communication, "some, but not all" (Begg and Harris, 1982; Newstead, 1995; Politzer, 1986; Sells, 1936). The strength of this bias towards pragmatic interpretations of the "some" quantifier is perhaps understandable given that most of the reasoning which people carry out on a daily basis is performed within thematic contexts with strong pragmatic associations. This phenomenon leads us naturally to the empirical question of whether reasoners are less liable to adopt pragmatic interpretations of the "some" quantifier when reasoning about material devoid of all meaningful relation.

The view that formal logic abstracts away all extraneous details and allows reasoners to concentrate solely on the underlying logical form of arguments derives from Kantian philosophy (Kant, in Smith, 1993). Based on this assumption, one might expect the likelihood that people reason logically to increase when a task is expressed in abstract, as opposed to thematic, content. However, natural language based studies of conditional reasoning suggest that performance actually improves as logical tasks become less abstract (Dominowski, 1995; Gilhooly and Falconer, 1974; van Duyne, 1974; Wason and Shapiro, 1971). It is claimed that the apparent facilitatory effects of thematic content can, in some cases, be explained by the possibility that reasoners simply read off responses from memory without performing the kind of logical analysis appropriate to the task (Griggs and Cox, 1982). These findings suggest that, despite the purely logical requirements of laboratory based deductive tasks, the semantic associations of thematic content can encourage reasoners to use incompatible, non-logical reasoning heuristics.

Previous studies of syllogistic reasoning also suggest that tasks whose thematic content eludes close associations with information stored in reasoners' semantic memories are more likely to incite responses that accord with prior belief, albeit sometimes at the expense of logical necessity (Barston, 1986; Henle and Michael, 1956; Janis and Frick, 1943; Thistlethwaite, 1950). More specifically, it is argued that the introduction of meaningful content into the syllogistic task can either facilitate or impair reasoning performance depending upon whether the inferences lead to conclusions which agree or disagree with prior beliefs respectively (Morgan and Morton, 1944). One study of syllogistic reasoning in thematic material (Wilkins, 1928), for example, suggests that reasoners fail to accept the logical task and respond solely according to their vague, intuitive feelings towards the plausibility of the relations between terms in the given premisses and the extent to which their putative conclusions concur with prior belief.

The theory of "belief bias" argues that an individual's personal convictions can exert a dominating influence on their judgements by causing him or her to accept arguments which are only plausible, rather than being logically valid. Several studies have shown that reasoners are liable to accept at face value arguments whose conclusions they believe regardless of their logical validity, but closely scrutinise those arguments whose conclusions do not conform with their prior beliefs (Barston, 1986; Evans et al., 1983; Janis and Frick, 1943; Revlis, 1975a). In general, the belief bias effects tend to become more evident as the content of a logical task becomes more closely related to the personal beliefs of the reasoner because firmly held convictions are likely to be held in spite of evidence against them (Morgan and Morton, 1944). According to Oakhill et al. (1990), there are three possible ways in which personal beliefs can affect syllogistic reasoning. Firstly, they can interfere with the interpretation of a syllogism's premisses, since a premiss may be illicitly converted or an illicit conversion blocked. Secondly, they can influence the deductive process by limiting the number of hypothetical possibilities that are considered. Or, thirdly, they can lead a reasoner to refine the number of putative conclusions that are derived from the deductive process. That personal convictions can preclude logical thought in these ways should be of concern to those communities whose work practices and livelihoods depend upon unimpeded logical reasoning, such as mathematics and software engineering.

When an individual is presented with the syllogistic task, two systems appear to compete for control of his or her cognitive processes: a logical system and a belief-driven pragmatic system. Politzer (1990) argues that it is usually the linguistic context of premiss information that determines which system takes control. For instance, problems containing a high degree of social content should incite the individual to apply pragmatic laws based on their own prior beliefs and knowledge, whereas problems containing purely abstract content should encourage the application of logical rules resembling those of a formal deductive system. Moreover, previous studies (Revlis and Leirer, 1980; Evans et al., 1983) suggest that reasoning performance deteriorates as logic and pragmatic beliefs point towards different conclusions, but improves as they concur.

Despite having no logical bearing on the syllogistic task, the possible psychological repercussions of manipulating term and premiss order has been the focus of much empirical concern. Although there appears to be general agreement that changing premiss order alone does not influence reasoning performance significantly (Dickstein, 1975; Wetherick and Gilhooly, 1990), several studies do suggest that the order of terms within premisses is significant (Begg and Harris, 1982). Johnson-Laird and Steedman (1978), for example, report a possible correlation between the types of conclusion drawn and syllogistic figure. The theory of "figural bias" claims that syllogistic figure determines the order in which reasoners relate end terms during information integration, and that a directional bias in their mental processes

makes it easier to scan the represented premisses in certain directions (Johnson-Laird and Bara, 1984). However, this theory applies only to premisses in which the reasoner is free to choose either non-middle term as the subject of the conclusion. For example, given the premiss pair "Some of the parents are scientists" and "All of the scientists are drivers", reasoners tend to draw the conclusion "Some of the parents are drivers" rather than the equally valid conclusion "Some of the drivers are parents." So, while certain premisses of the form S-M, M-P lead to valid conclusions of the form S-P, reversing the order of these same premisses to M-S, P-M can lead to valid conclusions of the form P-S which, despite contravening the Aristotelian convention of only drawing S-P conclusions, is logically sanctionable. The observation that reasoners do not draw these forms of conclusion with equal regularity appears to constitute evidence against Revlis' (1975a) theory that reasoners routinely convert premiss information.

Set-theoretic models of performance under the syllogistic task argue that people reason about syllogisms in ways analogous to those in which set-theoretic tools, such as Venn Diagrams or Euler Circles, are used in mathematics (Adams, 1984; Ceraso and Provitera, 1971; Erickson, 1974; 1978; Traub, 1977; Wetherick, 1993). A central tenet of this argument is that the interpretation of premiss information involves creating a combined mental representation of premiss information which shows the set relations that may exist between syllogistic terms. In order that the correct conclusion may be deduced, every possible combination of set relation which follows from the given premiss pair must be explored, and it is predicted that the difficulty of a syllogism will increase along with the number of ways in which its premisses can be represented (Ceraso and Provitera, 1971; Dickstein, 1978b; Erickson, 1974). Non-logical errors then become explainable as a consequence of reasoners' use of inappropriate representations or their failure to consider all possible hypothetical combinations. The possible Euler representations that are consistent with each form of syllogistic premiss are shown in Figure 5 (adapted from Erickson, 1974, p.310; Evans et al, 1993, p.220). That reasoners often seem inclined to consider only a limited number of the possible combined premiss representations is perhaps understandable given the mental effort that it would require to consider the entire set and the demanding nature of the syllogistic task.

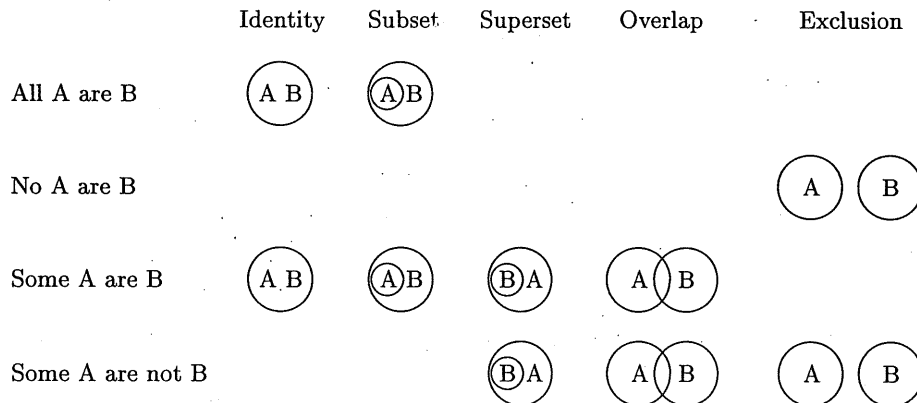


Figure 5: Possible Euler representations of syllogistic predicates

Another bias that is commonly cited in the context of the syllogistic task is the seemingly strong predisposition that reasoners have towards drawing propositional conclusions, despite the fact that most of the premisses presented in syllogistic studies tend to lead to indeterminate conclusions (Dickstein, 1975; 1976; Revlis, 1975a). It is argued that "propositional response bias" can either mislead reasoners into interpreting or combining premisses in ways that can only lead to determinate

conclusions, or by causing them to discount hypothetical possibilities that lead to indeterminate conclusions (Dickstein, 1978b). In other words, reasoners would generally prefer to draw a propositional conclusion rather than say that nothing follows from a given premiss pair. Experimenters often attribute this response bias to the disproportionate numbers of determinate and indeterminate arguments in the syllogistic task; assuming 64 possible premiss combinations, as per Dickstein's studies, less than one third lead to determinate conclusions. Errors then become attributable to reasoners' expectations that a greater proportion of the given premiss pairs will lead to determinate conclusions and reasoners' predisposition to draw propositional conclusions when indeterminate ones are appropriate.

To summarise our review of the literature, numerous independent studies of categorical syllogistic reasoning suggest that participants' errors are, in general, attributable to their application of similar processes which govern communication and reasoning with quantified statements in everyday linguistic experience. The fact that these studies point to the possible existence of encoding, processing and response biases suggests that the psychological causes of error in quantified reasoning are deep-rooted and unlikely to be overcome with quick-fix psychological remedies. This should be of serious concern to the software community in light of the fact that logically equivalent versions of the same quantified statements which have been shown to incite these errors and biases may be expressed in most, if not all, formal specification notations, and the fact that inaccurate reasoning about formal specifications can have serious detrimental repercussions on software projects and the quality of the products developed. The main aim of the present study was, then, to determine whether the users of formal methods are liable to succumb to the same forms of error and bias. As formal notations are not governed by the same linguistic principles which govern everyday communication in natural language, it appears unthinkable that the same non-logical tendencies could transfer over into the formal domain, especially when the syllogistic tasks are presented explicitly in formal logic and all participants have the appropriate logical training.

## 2 Experimental Methodology

*Participants.* A total of forty computing scientists volunteered to take part in the experiment comprising staff and students from academic universities and computing professionals from industrial software companies. All participants were recruited by personal invitation. These were divided equally 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 15 staff, 3 students and 2 software professionals. Their mean age was 34.65 years ( $s = 8.79$ ), ranging from 21 to 54, and all had studied at least one system of mathematical logic beforehand, such as the propositional or predicate calculi, or Boolean algebra. Their mean level of Z experience was 5.84 years ( $s = 4.55$ ). According to participants' personal ratings of Z expertise, the group comprised 8 expert, 11 proficient and 1 novice users of the Z notation. The TFL group comprised 13 staff, 1 student and 6 software professionals. Their mean age was 33.25 years ( $s = 9.79$ ), ranging from 23 to 66, and all had studied at least one system of mathematical logic beforehand. Their mean level of Z experience was 4.43 years ( $s = 3.89$ ), and the group comprised 5 expert, 10 proficient and 5 novice users.

*Design.* The study had a four factor mixed design. The first, between groups, factor was the degree of realistic material presented, abstract or thematic, corresponding to the two participant groups: AFL and TFL. The second, repeated measures, factor was the mood of the syllogistic premisses and had four levels for each premiss: A, E,

I and O, corresponding to the four moods shown in Figure 1. The third, repeated measures, factor was the syllogistic figure and had four levels, 1 to 4, corresponding to the four figures shown in Figure 2. As each syllogism comprises two premisses, a systematic variation of premiss mood effectively results in sixteen levels for this factor. A fourth, repeated measures, factor which applied only to the TFL group was the believability of the conclusion to be inferred and had two levels: intuitive and counter-intuitive.

The precise number of different premisses that may exist, and the number of valid conclusions that may be drawn, in the syllogistic task depends upon the criteria used. Some theorists argue that syllogisms in which the order of the premisses are simply reversed should count as an additional kind (Evans et al., 1993). However, premiss order has been shown not to affect reasoning performance significantly (Dickstein, 1975; Wetherick and Gilhooly, 1990). This factor was accordingly excluded from the present study. Normally, this design would yield 64 abstract tasks and 128 thematic tasks (16 moods  $\times$  4 figures  $\times$  2 levels of believability). Discounting the believability factor, the same design assumptions are made in a previous natural language based study of syllogistic reasoning (Dickstein, 1978a). However, for the practical purposes of this study, it was necessary to limit the number of tasks presented to participants. This was achieved by including only a representative sample of these syllogisms. The abstract tasks comprised 30 syllogisms (15 with determinate and 15 with indeterminate conclusions), whilst the thematic tasks comprised 40 syllogisms (15 with determinate believable conclusions, 15 with indeterminate believable conclusions, 5 with determinate unbelievable conclusions, and 5 with indeterminate unbelievable conclusions). Table 1 shows the forms of syllogism presented during the present study.

TABLE 1  
Structures of the syllogisms presented in the present study

Prem.	Conc.	Prem.	Conc.	Prem.	Conc.	Prem.	Conc.	Prem.	Conc.
AA1	A (I)	AA2	N	AA3	N	AA4	N	AI1*	I
AI3*	I	AO2	O	AO4*	N	AE2	E (O)	AE4	E (O)
IA3	I	IA4*	I	II3*	N	II4	N	IE1*	N
IE2*	N	IE4	N	OA1*	N	OA3*	O	OO3	N
OO4	N	EA1	E (O)	EA2	E (O)	EA3	N	EA4	N
EI1	O	EI2	O	EI3	O	EI4*	O	EE4	N

*Note:* Two versions of those syllogisms marked with an asterisk are presented to the TFL group; one with a believable conclusion, one with an unbelievable conclusion. Weak conclusions are given in parentheses.

In Dickstein's (1978a) study, where systematic variation of sixteen moods and four figures yields 64 premiss combinations, this gives rise to only 19 possible determinate conclusions.<sup>2</sup> However, owing to strong typing imposed by the notation in which the formalised tasks were expressed, four of the premiss pairs which normally lead to determinate conclusions led to indeterminate conclusions in the present study.<sup>3</sup> Although it would have been possible to overcome these restrictions and achieve a design that would allow for the same number of determinate conclusions to be drawn as in previous natural language based studies, this would not have been possible without compromising the complexity or the consistency of the manner in which the tasks were presented, either of which would have jeopardised the practicality of the study. It should be noted that this design gives rise to a lower ratio

of determinate to indeterminate syllogisms (approximately 2:1) than has been used in previous syllogistic studies (which has often exceeded 3:1). This is noteworthy because it is suggested that a bias towards propositional responses may be introduced as a consequence of the imbalance between determinate and indeterminate syllogisms (Chapman and Chapman, 1959; Dickstein, 1975; 1976; Revlis, 1975a; Traub, 1977). A more balanced design in this respect has favourable implications for achieving unbiased responses.

*Materials.* There were several reasons behind the decision to use the Z notation (Spivey, 1992) as the language in which to express the experimental tasks: it is popular in both academia and industry (Dean and Hinchey, 1996), it is claimed to be one of the more readable notations (Bowen, 1988), and its commercial viability has been demonstrated (Barden et al., 1992). In addition, Z is a strongly typed notation whose foundations lie in the propositional and predicate calculi (Lemmon, 1993) and set theory (Johnstone, 1987). Its underlying foundations are, then, representative of those underlying numerous other formal notations.

In order to simplify the formalisation of the syllogistic task a methodological approach was used to translate natural language based categorical syllogisms into equivalent forms in Z. In this respect, two main obstacles had to be overcome. Firstly, it was necessary to find formal operators that corresponded to the equivalent natural language quantifiers and qualifiers without altering the logical structure of the original task. Secondly, so as to avoid violating Z's strong type checking system, it was necessary to assign appropriate types to the variables, or "terms", of the formalised syllogism. Figure 6 shows the method used to translate the four possible forms of natural language based premiss into a logically equivalent Z form.

All A are B	$\forall x : Type \bullet A(x) \Rightarrow B(x)$
Some A are B	$\exists x : Type \bullet A(x) \wedge B(x)$
Some A are not B	$\exists x : Type \bullet A(x) \wedge \neg B(x)$
No A are B	$\neg \exists x : Type \bullet A(x) \wedge B(x)$

Figure 6: Z translations of the four syllogistic predicates

In order to facilitate the recall of relevant information from memory, meaningful identifiers were used for function names in the thematic versions of the tasks. These names were chosen so as to refer to concepts with which participants would be familiar, including: social groups, occupations, animals, foods and materials. In contrast, so as to minimise the possibility of interference from prior knowledge for the abstract tasks, arbitrary single letter identifiers were used in place of meaningful function names. The experiment's materials are exemplified by the abstract and thematic versions of the AA1 syllogistic task shown in Figures 7 and 8 respectively.

<sup>2</sup>Again, the precise number of determinate conclusions that may be drawn depends on the criteria used. In contrast, Johnson-Laird and Bara (1984) claim the existence of 27 valid determinate conclusions, but this view rests on the unconventional assumption that conclusions may be drawn in both directions: from S to P, and from P to S.

<sup>3</sup>The four premiss combinations AA3, AA4, EA3 and EA4 lead to determinate conclusions when expressed in natural language, yet lead to indeterminate conclusions in the present study. This is due to the fact that the terms of the formalised syllogisms must be assigned Z types, which are effectively mathematical sets, and that any two categorical premisses cannot give rise to a particular conclusion when the premiss terms might be assigned to empty sets. The following example contrasts natural language and formalised versions of an EA3 syllogism. The conclusion in the formalised version is indeterminate because the possibility that  $Food = \emptyset$  acts as a counter example to any possible determinate conclusion.

No oranges are apples	$\neg \exists f : Food \bullet orange(f) \wedge apple(f)$
All apples are fruits	$\forall f : Food \bullet apple(f) \Rightarrow fruit(f)$
Some fruits are not oranges	No determinate conclusion

$$\begin{aligned} \forall x : X \bullet B(x) \Rightarrow C(x) \\ \forall x : X \bullet A(x) \Rightarrow B(x) \end{aligned}$$

- (a)  $\exists x : X \bullet A(x) \wedge C(x)$
- (b)  $\forall x : X \bullet A(x) \Rightarrow C(x)$
- (c)  $\neg \exists x : X \bullet A(x) \wedge C(x)$
- (d) No valid conclusion

Figure 7: Abstract AA1 task

$$\begin{aligned} \forall p : Person \bullet human(p) \Rightarrow mortal(p) \\ \forall p : Person \bullet Greek(p) \Rightarrow human(p) \end{aligned}$$

- (a)  $\exists p : Person \bullet Greek(p) \wedge mortal(p)$
- (b)  $\forall p : Person \bullet Greek(p) \Rightarrow mortal(p)$
- (c)  $\neg \exists p : Person \bullet Greek(p) \wedge mortal(p)$
- (d) No valid conclusion

Figure 8: Thematic AA1 task

In seeking to investigate the effects of reasoners' personal beliefs on their reasoning performances, five thematic tasks were designed to lead to believable conclusions and five logically equivalent tasks were designed to lead to unbelievable conclusions. The nature of these tasks is exemplified in Figure 9, which requires the reasoner to draw the believable conclusion corresponding to "No rich people are poor", and Figure 10, which requires the reasoner to draw the unbelievable conclusion "Some communists are capitalists".

$$\begin{aligned} \neg \exists p : Person \bullet millionaire(p) \wedge poor(p) \\ \forall p : Person \bullet rich(p) \Rightarrow millionaire(p) \end{aligned}$$

$$\neg \exists p : Person \bullet rich(p) \wedge poor(p)$$

Figure 9: Believable EA1 task

$$\begin{aligned} \exists p : Person \bullet capitalist(p) \wedge Russian(p) \\ \forall p : Person \bullet Russian(p) \Rightarrow communist(p) \end{aligned}$$

$$\exists p : Person \bullet communist(p) \wedge capitalist(p)$$

Figure 10: Unbelievable IA4 task

*Procedure.* Prior to completing the main tasks, participants were asked to show their understandings of the four possible forms of quantified formal expression by completing four corresponding background tasks. Each task prompted the participant to select the closest natural English translation of a given formal Z expression corresponding to each of the four possible forms of syllogistic expression: A, E, I or O. It was hoped that these tasks would help to explain some of the erroneous trends that might arise in participants' responses during the main syllogistic tasks. Participants were then given the following instructions.

"In each of the tasks that follow, you will be shown two Z predicate expressions taken from an operational schema. You may assume that all of the named functions have been defined. 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 then 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 around one hour to complete."

For each main task, participants were shown two formal expressions representing the premisses of a categorical syllogism and three predicates representing the possible conclusions for determinate syllogisms, labelled "(a)" to "(c)". A fourth option was also presented, "No valid conclusion", which represented the conclusion for an indeterminate syllogism. This option was labelled "(d)" throughout. Participants were required to select the one conclusion which followed from the information contained in the premisses by circling the appropriate letter. 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.

Confidence rating:  Not confident  Guess  Confident

Prior to the experiment, participants were asked to provide the following biographical information: 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). All task sheets were computer generated. These were distributed to participants and completed anonymously then mailed back to the experimenter. All participants were tested on an individual basis.

### 3 Results

Comparison of the mean correctness scores for the two linguistic groups revealed that the AFL group (93.17% correct,  $\bar{x} = 27.95$  out of 30,  $s = 9.56$ ) outperformed the TFL group (90.38% correct,  $\bar{x} = 36.15$  out of 40,  $s = 12.63$ ). An analysis of variance revealed no significant differences between linguistic group performance at an overall level. Table 2 shows the frequencies of syllogisms solved correctly by the two linguistic groups.

TABLE 2  
Frequencies of syllogisms solved correctly by the two linguistic groups

AFL Group ( $n = 20$ )					TFL Group ( $n = 20$ )						
AA1	20 (0)	AA2	20	AA3	18	AA1	20 (0)	AA2	19	AA3	18
AA4	17	AI1	20	AI3	20	AA4	17	AI1	20	AI3	18
AO2	17	AO4	19	AE2	18 (0)	AO2	14	AO4	19	AE2	16 (2)
AE4	17 (1)	IA3	20	IA4	19	AE4	19 (1)	IA3	19	IA4	17
II3	20	II4	20	IE1	19	II3	19	II4	19	IE1	19
IE2	20	IE4	20	OA1	16	IE2	15	IE4	17	OA1	19
OA3	17	OO3	20	OO4	19	OA3	19	OO3	19	OO4	19
EA1	20 (0)	EA2	20 (0)	EA3	16	EA1	16 (1)	EA2	19 (1)	EA3	14
EA4	16	EI1	18	EI2	19	EA4	16	EI1	18	EI2	19
EI3	17	EI4	17	EE4	19	EI3	15	EI4	18	EE4	19

*Note:* Frequencies shown for the thematic group are for syllogisms with believable conclusions only. Weak conclusions are given in parentheses.

Table 3 shows the probabilities of inferring correct conclusions for particular classes of syllogism, as derived from each group's performance, and the statistical significance of each class of syllogism on participants' correctness, as derived from a series of analyses of variance. The probability values reveal that participants performed best with abstract premisses which were either both affirmative or both negative, and worst with thematic premisses in the second or third syllogistic figure. The analyses of variance reveal that participants' levels of correctness were significantly affected by syllogisms with determinate and indeterminate conclusions, matching and unmatching premiss moods, two affirmative but not two negative premiss polarities, second, third and fourth figures, and an effect approaching significance for first figure syllogisms.



TABLE 3  
Probabilities of drawing correct conclusions ( $0 \leq p \leq 1$ ) and statistical significance

Syllogism	<i>N</i>	AFL Group	TFL Group	Significance
Determinate	15	0.93 (0.25)	0.91 (0.29)	$F_{(14)} = 2.79, p < 0.01$
Indeterminate	15	0.93 (0.26)	0.89 (0.31)	$F_{(14)} = 3.75, p < 0.01$
Matching Mood	9	0.96 (0.19)	0.94 (0.24)	$F_{(8)} = 3.27, p < 0.01$
Unmatching Mood	21	0.92 (0.27)	0.88 (0.32)	$F_{(20)} = 2.55, p < 0.01$
Affirmative Mood	10	0.97 (0.17)	0.93 (0.28)	$F_{(9)} = 2.86, p = 0.03$
Negative Mood	3	0.97 (0.18)	0.95 (0.22)	$F_{(2)} = 0.49, p = 0.62$
First Figure	6	0.94 (0.24)	0.94 (0.24)	$F_{(5)} = 1.96, p = 0.09$
Second Figure	6	0.95 (0.22)	0.88 (0.33)	$F_{(5)} = 3.73, p < 0.01$
Third Figure	8	0.93 (0.26)	0.88 (0.33)	$F_{(7)} = 4.09, p < 0.01$
Fourth Figure	10	0.92 (0.27)	0.91 (0.29)	$F_{(9)} = 1.89, p = 0.05$

Note: Standard deviations are shown in parentheses.

Judging by the confidence ratings shown in Table 4, participants were highly confident in the correctness of their responses throughout all tasks. With regards to syllogistic determinacy, an analysis of variance revealed no significant effects of determinate or indeterminate syllogisms on confidence. With regards to similarity of syllogistic mood, an analysis of variance revealed no significant effects of syllogisms with matching quantifiers on confidence, however, it did reveal a significant effect of syllogisms with unmatching quantifiers on confidence ( $F_{(20)} = 1.75, p = 0.02$ ). With regards to polarity of syllogistic mood, an analysis of variance revealed an effect approaching significance for syllogisms with two affirmative premisses on confidence ( $F_{(9)} = 1.85, p = 0.06$ ), but no significant effect of syllogisms with two negative premisses on confidence. With regards to syllogistic figure, an analysis of variance revealed a significant effect of first figure syllogisms on confidence ( $F_{(5)} = 2.59, p = 0.02$ ), but no significant effects of second, third or fourth figure syllogisms on confidence. The analyses failed to reveal any significant effects of linguistic group type on participants' confidence.

TABLE 4  
Mean confidence ratings for the believable syllogisms

AFL Group ( <i>n</i> = 20)					TFL Group ( <i>n</i> = 20)						
AA1	2.85	AA2	2.75	AA3	2.75	AA1	3.00	AA2	2.95	AA3	2.95
AA4	2.70	AI1	2.90	AI3	2.80	AA4	2.85	AI1	3.00	AI3	2.95
AO2	2.85	AO4	2.75	AE2	2.75	AO2	3.00	AO4	2.85	AE2	3.00
AE4	2.75	IA3	2.85	IA4	2.80	AE4	2.90	IA3	3.00	IA4	2.95
II3	2.70	II4	2.75	IE1	2.75	II3	3.00	II4	2.95	IE1	2.80
IE2	2.75	IE4	2.70	OA1	2.75	IE2	2.95	IE4	2.95	OA1	2.85
OA3	2.80	OO3	2.75	OO4	2.75	OA3	3.00	OO3	2.90	OO4	2.90
EA1	2.75	EA2	2.85	EA3	2.85	EA1	3.00	EA2	3.00	EA3	3.00
EA4	2.70	EI1	2.80	EI2	2.85	EA4	2.85	EI1	3.00	EI2	2.90
EI3	2.85	EI4	2.85	EE4	2.85	EI3	2.90	EI4	3.00	EE4	2.85

Note: All confidence ratings range from 1.00 (not confident) to 3.00 (confident). Ratings shown for the thematic group are for syllogisms with believable conclusions only.

Table 5 shows the frequencies of syllogisms with unbelievable conclusions solved correctly by the thematic group and their corresponding confidence ratings. An analysis of variance revealed a significant effect of unbelievable syllogism type on participants' correctness ( $F_{(9)} = 1.97, p = 0.05$ ), but no significant effect of unbelievable syllogism type on participants' confidence.

TABLE 5  
Frequencies of syllogisms with unbelievable conclusions solved ( $n = 20$ )

Syllogism	Correct	Mean CR	Syllogism	Correct	Mean CR
AI1	18	3.00	IE1	17	2.85
AI3	19	2.95	IE2	18	2.85
AO4	18	2.95	OA1	19	2.90
IA4	20	2.95	OA3	20	3.00
II3	19	2.90	EI4	15	2.90

Table 6 shows the four formal expressions presented to participants as part of the experiment's background tasks and the natural language translations offered to participants as the response options in each case. A series of chi-square tests reveal that participants' response selections were significant in all cases: "All" ( $\chi^2_{(3)} = 104.8, p < 0.01$ ), "Some" ( $\chi^2_{(3)} = 52.6, p < 0.01$ ), "Some ... not" ( $\chi^2_{(3)} = 52.6, p < 0.01$ ), and "None" ( $\chi^2_{(3)} = 120, p < 0.01$ ). Inspection of Table 6 reveals that nearly all participants succeeded in selecting natural language translations corresponding to unambiguous set-theoretic interpretations of the two universal expressions, quantified by " $\forall$ " (All) and " $\neg\exists$ " (None), but nearly three quarters of participants failed to select unambiguous set-theoretic translations of the two particular expressions, quantified by " $\exists$ " (Some) and " $\exists \dots \neg$ " (Some ... not).

TABLE 6  
Frequencies of participants' selections during the background tasks ( $N = 40$ )

	$\forall t : T \bullet A(t) \Rightarrow B(t)$		
All As are Bs*	38	At least one (possibly all) As are Bs	0
Possibly all As are Bs	2	Some As are Bs	0
	$\exists t : T \bullet A(t) \wedge B(t)$		
At least one A is a B	29	At least one (possibly all) As are Bs*	9
Exactly one A is a B	0	Some As are Bs	2
	$\exists t : T \bullet A(t) \wedge \neg B(t)$		
At least one A is not a B	29	At least one (possibly all) As are not Bs*	9
Exactly one A is not a B	0	Some As are not Bs	2
	$\neg\exists t : T \bullet A(t) \wedge B(t)$		
None of the As are Bs*	40	At least one (possibly none) of the As are Bs	0
Possibly none of the As are Bs	0	Exactly one A is not a B	0

Note: Unambiguous set-theoretic translations are indicated with an asterisk.

Analysis by linear regression failed to reveal any significant correlations between participants' levels of correctness and either their ratings of expertise, their levels of experience or their ages, in either linguistic group. These findings suggest that participants' increased experience or expertise with the Z notation were not primarily responsible for their levels of performance.

#### 4 Discussion

Although the overall differences in group performance were not significant at a statistical level, evidence that the abstract group outperformed the thematic group is reflected in: the equal or lower rates of correctness observed for all of the main classes of syllogism (as shown in Table 3), the mean rates of overall correctness observed for the two groups (93.17% for the abstract group versus 90.38% for the thematic group) and the fact that the abstract group achieved three times as many perfect scores for individual syllogisms than the thematic group. That participants exhibited signs of improved reasoning under the abstract condition is supported by previous natural language based studies of syllogistic reasoning. Wilkins (1928, p.77), for instance, attributes improved syllogistic performance under the abstract condition to the "bad habits of everyday reasoning which are much in force in the familiar situation, but are not so influential when the material is symbolic or unfamiliar". The fact that similar overall mean rates of correctness were observed for the two conditions is also supported in the cognitive literature. Henle and Michael (1956), for example, suggest that when reasoners' beliefs are not held with a sufficient degree of conviction, or reasoners are indifferent to the real world relations established by the task, they are unlikely to cause major distortions of logical reasoning. In such cases, then, performance is likely to be similar under both abstract and thematic conditions. The fact that more sporadic rates of correctness were observed in the thematic group suggests that the presence of meaningful content affected performance in certain tasks but not in others. This is supported by cognitive studies which suggest that any facilitative or depreciative effects caused by changed material are entirely specific to the task and the extent to which its content relates to the reasoner's prior beliefs (Barston, 1986; Traub, 1977).

Despite their relatively high confidence ratings, participants' apparent susceptibility to logical error suggests that many were overconfident in the correctness of their responses, particularly in the thematic group. An interesting correlation between correctness and confidence is evident in the results. Although the abstract group outperformed the thematic group overall, the mean confidence ratings for every thematic task is higher than the corresponding rating for its abstract counterpart, with only one exception; namely task EE4, where the two confidence ratings are equal. In general, one might expect a reasoner's confidence to increase along with their appreciation of a task's requirements and, consequently, with their levels of correctness. However, given that the abstract group outperformed the thematic group overall, the observed results actually run contrary to this expectation. This trend may be attributable to the recognition of familiar everyday terms which led the thematic group to believe that non-logical everyday reasoning heuristics were sufficient for the tasks at hand, and the recognition of purely symbolic terms which led the abstract group to believe that a logical approach was more appropriate. Since the use of non-logical heuristics is generally perceived to involve a less complicated or mentally intensive analysis as that required by a purely logical approach, this may account for the observed differences in group confidence.

## 4.1 Signs of Classical Error and Bias

“A person is likely to accept a conclusion which expresses his convictions with little regard for the correctness or incorrectness of the inferences involved . . . the only circumstance under which we can be relatively sure that the inferences of a person will be logical is when they lead to a conclusion that has already been accepted” (Morgan and Morton, 1944, p.39).

The theory of belief bias claims that, rather than use logical analysis, people are liable to respond according to their personal convictions towards a task's meaningful content (Barston, 1986; Begg and Harris, 1982; Janis and Frick, 1943; Revlin and Leirer, 1980). More specifically, it is hypothesised that people will tend to accept conclusions which they believe, and reject conclusions which they disbelieve, with little regard for their logical relation to the given premisses. Previous syllogistic studies vary in their degree of support for this hypothesis and report rates of correctness which seem dependent on whether the logic of the task and individuals' prior beliefs converge on the same conclusion. Evans et al., (1983), for example, report overall rates of correctness as high as 97% when logic accords with belief and as low as 27% when logic conflicts with belief, while Revlin et al. (1980) report respective rates of 83% and 67%. These notable differences in performance lead the experimenters to speculate whether reasoners are liable to accept believable conclusions uncritically and only resort to logical analysis when they encounter premisses which lead to unbelievable conclusions; a trend which has been documented as the theory of “selective scrutiny” (Barston, 1986; Evans et al., 1993). However, this is not a view which is shared universally in the cognitive literature. Other studies report no discernible differences in correctness for those syllogisms phrased completely in abstract material and those leading to thematic conclusions which concur with popular belief (Henle and Michael, 1956; Newstead, 1995).

Inspection of the mean scores for those ten syllogisms with both believable and unbelievable conclusions reveals that the effects of changed material were not as marked as has been reported in previous natural language based studies: 89% where conclusions agreed with belief against 86% where conclusions conflicted with belief. Rather than demonstrating clear overall differences in performance under the two conditions, the sporadic rates of correctness suggest that any facilitation or suppression of reasoning performance was entirely task or content specific. Signs of performance facilitation, for example, is evident in the perfect scores observed for the thematic AA1 and AI1 syllogisms leading to believable conclusions. Here, it is possible that real world knowledge guided participants to the intuitively plausible, valid conclusions corresponding to the natural language statements “All Greeks are mortal” and “Some students are rewarded” respectively. However, the fact that perfect scores were also observed for these tasks in the abstract group suggests that the logic of these specific tasks was comparatively easy for participants to grasp in any case, and that this may have been the primary reason for participants' high rates of correctness. In the case of the IA4 and OA3 syllogisms leading to unbelievable conclusions, participants' prior convictions towards the task content did not appear sufficiently strong to lead them away from drawing the correct counter-intuitive conclusions, “Some communists are capitalists” and “Some athletes are not healthy”, respectively. However, the fact that similarly high rates of correctness were not observed for the other eight syllogisms leading to unbelievable conclusions appears to lend further credence to the belief bias hypothesis.

Besides signs of performance facilitation, the results of the present study also suggest that participants' selection of the logically correct responses were, in several cases, suppressed by their prior convictions towards the content of the thematic tasks. In the case of the IE2 syllogism, for example, no logically determinate conclusion can be drawn. The perfect response rate observed for the abstract version

of this task might be attributed to the use of single letter syllogistic terms which did not elicit misleading prior beliefs, but instead provided participants with an unimpeded view of the task's logical structure. In a logically equivalent thematic version of the same task, participants were presented with formalised premisses corresponding to the natural language expressions "Some scientists are methodical" and "No drunkards are methodical". Although most participants succeeded in giving indeterminate responses to these premisses, the fact that the erroneous determinate conclusion, "No drunkards are scientists", was endorsed by one quarter of participants might be attributed to the pragmatic implication that drunkards are too disorganised to be methodical. That is, participants' prior convictions towards the real world referents of the syllogistic terms appeared to distort their view of the logical task. But besides encouraging participants to suppress the selection of correct indeterminate responses in favour of incorrect determinate ones, belief bias also appears to have had the converse effect. This trend was most evident where participants preferred to give indeterminate conclusions rather than unbelievable ones. In the case of the EI4 syllogism, for example, one quarter of participants failed to give a logically valid response corresponding to "Some traitors are not disloyal", which is self-evidently counter-intuitive since traitorous people must, by definition, at some point have been disloyal.

Atmosphere theory predicts that, where an individual is unwilling or unable to use logical deduction to expose the link between the subject and predicate of the correct conclusion, he or she will be inclined to draw a conclusion similar in quantity or quality to the predicates of the given premisses. This hypothesis gives rise to several specific predictions: AA premisses will give rise to A conclusions, AE, EA or EE premisses will give rise to E conclusions, II, AI or IA premisses will give rise to I conclusions, and OO, AO, OA, EI or IE premisses will give rise to O conclusions (Sells, 1936; Simpson and Johnson, 1966; Woodworth and Sells, 1935). The results of the present study offer mixed support for these predictions. Insofar as perfect scores were observed for the abstract syllogisms AA1, EA1, AI1, IA3, and thematic syllogisms AA1, AI1, IA4 (unbelievable), OA3 (unbelievable), the results suggest that performance was facilitated where logic and atmosphere theory pointed to the same conclusion. On the other hand, there were several cases where many participants still failed to draw the correct conclusion, even where this was dictated both by logic and atmosphere theory; the abstract syllogisms AE4, AO2, EI3 and EI4, and thematic syllogisms EA1, AO2, EI3 and EI4 (unbelievable). For those remaining cases where logic and atmosphere theory pointed towards different conclusions, the high rates of observed correctness suggest that logic appeared to exert a much more dominating influence on participants' reasoning. Signs of atmosphere bias are, however, evident in the relatively high rates of non-logical conclusions given in response to the abstract syllogisms EA3 and OA1, and the thematic syllogism EA3, which are also consistent with the predictions of the cognitive literature.

These findings do suggest that atmosphere bias can lead reasoners away from what is logically sanctionable under formalised conditions. However, they do not concur with the extremely high error rates observed in previous natural language based studies. Sells and Koob (1937), for example, report instances in which atmosphere bias seemingly led to error rates exceeding 90%. These extremely high error rates, however, could be attributed to the severe time pressures imposed by the experimenters, which may have prevented participants from conducting a full logical analysis of the task. The fact that there was no time limit imposed during the present study may, then, be partly responsible for participants' higher rates of logical correctness.

There were several instances where neither logic or atmosphere theory appeared to account for the trends in participants' responses. Take, for example, the fact that one quarter of participants selected an E conclusion in response to the thematic IE2

sylllogism. It is possible that errors of this kind may be attributable to the influence of matching bias; that is, the tendency to select a conclusion whose quantifier matches one of those in the given premiss pair. Inspection of Table 3 suggests that participants found it easier to solve syllogisms with matching premiss moods than those without in both linguistic groups. Taken at face value, this overall comparison appears to lend credence to the matching bias hypothesis. But, given that nearly all of the syllogisms with matching premiss moods led to logical conclusions with different moods, this performance differential, taken together with the relative lack of support for atmosphere theory, suggests that formalisation of the syllogistic task actually reduces the likelihood that reasoners will respond according to vague, intuitive feelings towards the overall mood created by the given premisses.

Inspection of Tables 2 and 5 reveals that the lowest scores observed in both linguistic groups were for syllogisms with just one negative premiss mood. At least one fifth of participants erred on: the abstract syllogisms EA4, OA1 and OA3, the thematic syllogisms AO2, IE2, EA1, EA4, EI3, EA3, and EI4 (unbelievable). In comparison, participants' achieved higher rates of correctness on syllogisms with matching premiss moods, even when both were negative. This finding is supported by previous studies of disjunctive reasoning which suggest that premisses containing one negative component are actually more difficult than those containing two (Evans and Newstead, 1980; Roberge, 1976; 1978). As a possible explanation for this phenomenon, it is theorised that people implicitly drop the negatives when they appear in both terms because they find it easier to reason with affirmatives only. People's response patterns suggest that inferences involving two negatives are actually simpler than those containing just one, then, because the latter cannot be easily converted into an affirmative form. However, the conversion of premisses in this manner often does not lead to logically valid conclusions, as suggested by the lower rates of correctness observed for those syllogisms containing one negative premiss in the present study.

According to the theory of figural bias, syllogistic figure determines the ease with which reasoners relate end terms during premiss integration (Begg and Harris, 1982; Johnson-Laird and Steedman, 1978). The theory gives rise to the prediction that performance becomes less logical as participants' proceed from Figures 1 to 4 (Dickstein, 1978a), which is supported by the rates of overall correctness observed for determinate syllogisms in a previous natural language based study: 79% in figure 1, 76% in figure 2, 59% in figure 3, and 49% in figure 4 (Erickson, 1974). Table 7 shows a comparison of participants' performance in the present study under each of the four syllogistic figures for both determinate and indeterminate syllogisms.

TABLE 7  
Percentages of correct conclusions drawn in the four syllogistic figures

Group	Figure 1		Figure 2		Figure 3		Figure 4	
	Valid	Invalid	Valid	Invalid	Valid	Invalid	Valid	Invalid
AFL	97.50	87.50	92.50	100.00	92.50	92.50	90.00	92.86
TFL	91.25	92.50	85.00	87.50	93.13	91.25	88.75	90.00

Taking into account both groups and the thematic syllogisms with unbelievable conclusions, the overall rates of correctness for the four syllogistic figures were as follows: 93.00% in figure 1, 90.00% in figure 2, 91.32% in figure 3, and 90.65% in figure 4. The fact that these rates are considerably higher and more equally balanced in comparison to previous natural language based studies suggests that

sylogistic figure rarely, if at all, influenced participants' reasoning during the present formalised study. Given that performance was most logical in the first figure, the results do, however, lend some support to the hypothesis that logical performance should be facilitated on first figure syllogisms (Dickstein, 1978a; Johnson-Laird and Bara, 1984), where the correct conclusion can be exposed simply by scanning the given premisses in a forwards direction.

The frequent occurrence of the "Same M" fallacy in natural language based syllogistic studies has, in general, been attributed to participants' tendency to adhere to the Gricean maxim of relation. That is, the assumption that experimenters would not intentionally make two completely unrelated statements in sequence leads them to believe that the same class members of the middle term are being referred to in both premisses. Previous studies suggest that the effects of this fallacy are most evident in participants' responses to II syllogisms, where the middle terms seemingly share the common property of being related to both end terms, and OO syllogisms, where the middle terms seemingly share the common property of being unrelated to both end terms (Chapman and Chapman, 1959; Dickstein, 1975; 1976; Haviland and Clark, 1975). The fact that only one participant gave responses consistent with these findings for each of the II4, OO3 and OO4 tasks, and that nearly all others gave indeterminate responses, appears to provide limited support for this hypothesis. However, the fact that up to one quarter of participants endorsed determinate conclusions in response to numerous indeterminate tasks suggests that participants believed that some form of logical relation existed between the given premiss pairs, and, in order that a logical relation may exist between any two syllogistic premisses it is an essential pre-requisite that they share a middle term which necessarily refers to the same class members. In this sense, the results observed for the abstract syllogisms AA4, EA3, EA4 and OA1, and thematic syllogisms AA4, IE2, IE4, EA3 and EA4, support the hypothesis that the experimenters were perceived by participants, even in explicitly logical contexts, as conforming with Gricean conventions of ordinary discourse.

Previously, the design of syllogistic studies has tended to create an imbalance between the ratio of indeterminate to determinate tasks, often in excess of 3:1 (see for example: Chapman and Chapman, 1959; Dickstein, 1978a). The theory of propositional response bias claims that individuals are predisposed to believe that there will be a greater proportion of determinate syllogisms than there actually are, and that this expectation contributes to their downfall on indeterminate tasks (Revlis, 1975a). It would appear, then, that the effects of the bias may become more or less evident depending on the way in which the study is formulated. Previous studies confirm that performance is generally better for syllogisms with determinate rather than indeterminate conclusions (Dickstein, 1976; 1978b; Evans et al., 1983). For example, Roberge (1970) observes 51.2% correct responses to determinate syllogisms versus 35.8% correct responses to indeterminate syllogisms, whilst Dickstein (1975) observes 72.6% and 58.2% respectively. However, inspection of Table 3 reveals that only a slightly higher mean rate of correctness was observed for determinate syllogisms in the present study suggests that the greater proportion of determinate tasks presented (roughly 2:1) may have curbed somewhat participants' predisposition to give propositional responses.

The erroneous trends observed in previous studies of syllogistic reasoning (Begg and Denny, 1969; Chapman and Chapman, 1959) are ascribed by the experimenters jointly to the influence of illicit premiss conversion, where participants appear to accept the converse of universal affirmatives or of particular negatives, and probabilistic inference, where individuals appear to succumb to the "Same M" fallacy. The results give rise to several predictions for particular premiss pairs, some of which gain support from the results of the present study. The prediction of an E response for EA and IE premiss pairs is supported by the numbers of erroneous

E conclusions given in response to the abstract EA3, thematic EA3, thematic IE2 and IE4 tasks. Furthermore, the prediction of an O response for OA premiss pairs is supported by the number of erroneous O conclusions given in response to the abstract OA1 task. All of these predictions gain corroboration from the results of Dickstein (1978a) and Roberge (1970), who also report high rates of erroneous responses consistent with illicit conversion theory rather than the rules of logic for these specific syllogisms. However, none of the experimenters' remaining predictions with regard to illicit conversion and probabilistic bias gain support from the results of the present study. Notably, all of the syllogisms in which conversion appears to have led reasoners astray from the dictates of logic during the present study lead to indeterminate conclusions. It is possible, then, that many of participants' non-logical conversions of the given premisses were endorsed as a consequence of their bias towards propositional conclusions generally.

Illicit conversion of the A premiss in the indeterminate thematic EA4 task might explain why one quarter of participants endorsed determinate A responses. This is supported by empirical evidence which suggests that participants are inclined to convert "All A are B" to "All A are B and all B are A" (Dickstein, 1981) or "All B are A" (Politzer, 1990). In this case, conversion may have been facilitated by participants' social knowledge pertaining to the task, which appeared to almost invite an illicit conversion of the indeterminate EA4 task into a determinate EA1 task. The natural language forms of these tasks are illustrated in Figures 11 and 12 respectively. The slightly lower rate of E conclusions given in response to the abstract version of the same task, then, might be attributed to the fact that conversion of the A premiss was not facilitated by the intuitive plausibility of the relation between terms in the resulting conclusion.

No churchgoers are atheists  
 All churchgoers are devout people  
 -----  
 Nothing

Figure 11: Original EA3 task

No churchgoers are atheists  
 All devout people are churchgoers  
 -----  
 No devout people are atheists

Figure 12: Converted EA1 task

That one quarter of participants selected a non-logical E conclusion in response to the thematic IE2 task does not appear to be so easily explainable in terms of conversion theory. However, as a tentative explanation, it is postulated that these participants firstly converted the original second figure syllogism into the first figure by legitimately interchanging the terms in the I premiss, thereby establishing a forwards directed chain of transitive relations, which, according to previous studies (Dickstein, 1978a; Johnson-Laird and Bara, 1984), are easier for individuals to represent mentally. Once this more clearly visible transitive chain is combined with the realistic plausibility of the converted relations and the Gricean assumption that experimenters would not intentionally make two unconnected statements in succession, the individual is seemingly under obligation to draw a determinate conclusion, even where this is not logically sanctionable, as illustrated in Figures 13 and 14. Although it is postulated that conversion was not the sole determinant of participants' errors during this task, then, the conversion of P-M, S-M to M-P, S-M, or even S-M, M-P, does create a forwards directed chain of transitive relations which seems to encourage conclusions of the form S-P.

Some scientists are methodical  
 No drunkards are methodical  
 -----  
 Nothing

Figure 13: Original IE2 task

Some methodical people are scientists  
 No drunkards are methodical  
 -----  
 No drunkards are scientists

Figure 14: Converted IE1 task



As a tentative explanation for the three non-logical E responses to the thematic IE4 task, we propose that these may be attributed to participants' illicit conversion of the particular affirmative premiss into a universal affirmative, which effectively turned the indeterminate IE4 syllogism into a determinate AE4 syllogism, as illustrated in Figures 15 and 16. This form of conversion, documented as "conversion by substitution" by Politzer (1990), occurs when an individual simply replaces the original premiss with a completely new one. In this case, it seems that the bias towards propositional conclusions may have also been partly responsible for the observed errors since these participants seemed determined to draw a determinate conclusion, even where the content of the converted premiss and resulting conclusion ran contrary to popular belief.

Some edible foods are vegetables No vegetables are minerals <hr style="width: 80%; margin: 5px auto;"/> Nothing	All edible foods are vegetables No vegetables are minerals <hr style="width: 80%; margin: 5px auto;"/> No minerals are edible
---	---

Figure 15: Original IE4 task

Figure 16: Converted AE4 task

Previous studies suggest that the illicit conversion of universal affirmatives is responsible for a large proportion of errors in the syllogistic task (Dickstein, 1981; Newstead, 1989; Newstead and Griggs, 1983; Politzer, 1990). It is postulated here that this form of conversion led to the four non-logical A responses to the abstract OA1 task. The fact that high rates of correctness were observed for both thematic versions of the task suggests that the thematic statements created by conversion of the A premiss in these cases may have ran contrary to participants' prior beliefs and blocked their attempts to draw determinate conclusions. That conversion for these tasks would have led to the counter-intuitive assertions "All birds are owls" and "All mammals are dogs" lends support to this hypothesis. Evidence of illicit conversion blocked by the realistic associations of thematic material is reported throughout the cognitive literature (Ceraso and Provitera, 1971; Chapman and Chapman, 1959; Evans et al., 1983; Henle and Michael, 1956; Newstead, 1990; Revlis, 1975a; Revlin et al., 1980). Figures 17 and 18 illustrate how we propose that illicit conversion led participants to err in the abstract version of the same task, where the thematic connotations of the task material did not appear sufficiently strong to block conversion.

Some B are not C All A are B <hr style="width: 80%; margin: 5px auto;"/> Nothing	Some B are not C All B are A <hr style="width: 80%; margin: 5px auto;"/> Some A are not C
--	---

Figure 17: Original OA1 task

Figure 18: Converted OA3 task

The results of the present study reaffirm previous cognitive findings that conversion interacts with other cognitive heuristics and biases, such as probabilistic bias (Begg and Denny, 1969; Roberge, 1970), during the encoding of premiss information and response generation. Judging by the trends in the observed responses, the likelihood of a conversion being accepted appears to depend upon two factors. Firstly, it seems to depend upon the degree of perceivable symmetry that exists between the end terms of a syllogism. This was evident during the EA3 and IE2 tasks, where participants apparently converted the given premisses into the first syllogistic figure before attempting to generate putative conclusions. This is supported by the findings of previous studies (Begg and Harris, 1982; Dickstein, 1978a) which suggest that participants who are led to believe that the experimenter has not

presented a premiss pair in an ordered and symmetrical manner will be inclined to convert it to another syllogistic figure, where the relations between end terms of the conclusion are more readily apparent. Secondly, the likelihood of a conversion being accepted seems to depend not on the degree of thematic material used alone, but on whether the material establishes believable relations in converted premisses or putative conclusions, according to participants' conceptions of the real world. This was evident in the thematic EA3 task, where premiss conversion led to thematic relations which conformed with popular social beliefs, and in the OA1 task, where premiss conversion was seemingly blocked when it led to forms that contradicted popular zoological knowledge. Given that erroneous responses were attributable to illicit conversion in several isolated cases, and that participants' use of conversion was not evident outside of the aforementioned conditions, the results of the study are not, however, consistent with Revlin and Leirer's (1980) hypothesis that conversion is a routine part of reasoners' interpretation of task stimuli.

It is argued that a large proportion of participants' errors in the syllogistic task are attributable to the tendency to "treat logical statements as if they are obscure attempts at communication, and interpret them by the same conventions they would use in normal discourse" (Begg and Harris, 1982, p.596). Grice's (1975) seminal work on conversational implicature represents an attempt to explicate some of those conventions which govern the implicit and explicit use of language in everyday communication. In particular, previous studies (Newstead, 1989; Politzer, 1986) suggest that it is participants' adherence or non-adherence to the maxims of quantity and relation, as proposed by Grice in his Conversational Principle, which are primarily responsible for participants' non-logical responses.

Adherence to the Gricean maxim of relation requires that speakers keep the content of their communicative contributions as relevant as possible. Perhaps owing to the generality of the hypotheses which stem from it, the maxim of relation can explain a wide range of errors in the syllogistic task because participants are led to believe, by their failure to differentiate the laws of logic from the laws of ordinary discourse, that experimenters would not intentionally make two unrelated statements in such close succession without there existing some form of link between them. In the present study, the strong inclination to adhere to this maxim appears evident in the large number of determinate conclusions that were endorsed in response to indeterminate syllogisms, where it seemingly gave rise to a bias against indeterminate conclusions, and in the large number of invalid determinate conclusions drawn by participants because their adherence to the maxim misled them into believing that adjacent premisses were logically related by the same middle term. That the rates of correctness observed for the indeterminate tasks are, nevertheless, generally higher than have been reported in previous natural language based studies may be ascribed to the explicitly logical nature of the tasks and participants' experience of mathematical logic, where it is part of the accepted norm that two consecutive statements may be unrelated.

Adherence to the Gricean maxim of quantity requires that speakers make their communicative contributions as informative as is required for the purposes of the situation. Normally, in ordinary speech, it is assumed that speakers routinely abide by the maxim and will attempt to divulge as much useful information that they have access to; they will not say "some" when "all" is applicable, and they will not say "some ... not" when "no" is applicable. In ordinary speech, then, the particular quantifiers "some" is given the partitive interpretation, "at least one, but not all", and "some ... not" is given the partitive interpretation "at least one is not, but not none". In other words, people's shared pragmatic knowledge leads them to believe that the use of the I and O particular quantifiers in ordinary communication preclude the possibility of their universal counterparts (Newstead, 1989). Such knowledge, however, is incompatible with the use of the particular

quantifiers in logic, where "some" means "at least one, and possibly all", and "some ... not" means "at least one is not, and possibly none".

The quantifiers "some" and "some ... not", then, as they appear in most natural language studies of syllogistic reasoning are ambiguous because the individual is uncertain of whether to adopt an everyday (partitive) interpretation or a logical (partitive, but possibly universal) interpretation. When the syllogistic task is expressed purely in the symbology of formal logic, however, one might expect more participants to be cued into interpretations which conform with the dictates of logic rather than the conventions of everyday communication. Further support for this hypothesis is gained from the results of previous studies. Chapman and Chapman (1959), for example, propose that the qualifier "are" in syllogistic predicates encourages readers to assume an identity relation, "is equal to", between terms when an inclusion relation, "is included in", would be more appropriate from a set-theoretic point of view. The experimenters claim that this form of interpretation encourages individuals to make unwarranted assumptions of symmetry between terms, in a manner similar to that hypothesised by Tsal (1977). One might expect, however, that the substitution of formal operators with precise formal meanings for the supposedly ambiguous qualifiers "are" and "are not" would further help to dispell the ambiguities from the syllogistic task.

Inspection of participants' responses to the background tasks reveals that nearly all participants drew the most precise set-theoretic interpretations of the universal statements corresponding to A and E premisses. This suggests that participants generally adhered to the Gricean maxim of quantity, since they preferred to say "all" where "possibly all" and "some" were also possible, and "none" where "exactly one is not" and "possibly none" were possible. Only two participants appeared to adopt subset interpretations, "possibly all", of the A premiss, which, for these participants, seemed to imply that "Some As might not be Bs". Logically speaking, however, this is an unconditional categorical statement in which "All of the As are Bs" is definitively asserted, and the possibility that "Some of the As might not be Bs" simply cannot arise. This raises the question of why anyone would say "possibly all" when they know "all" to be the case? A possible explanation is that these participants adopted a more conservative form of interpretation, which can often be sufficient for the purposes of everyday reasoning but leads to error in strictly logical tasks.

Participants' willingness to conform with the Gricean maxim of quantity is also evident in the observed results to the main experimental tasks. Inspection of Table 2 reveals that a total of only six weak conclusions were endorsed where stronger versions were also possible. That participants appeared to exhibit such a degree of preference for universal, rather than particular, conclusions in the present formalised study provides little support for Woodworth and Sells' (1935) theory of "caution bias", which claims that reasoners are generally conservative estimators. However, the fact that all except one of these weak conclusions were chosen by participants from the thematic group suggests that the degree of meaningful content can be highly influential in determining the strength of conclusion endorsed by reasoners.

In contrast, the Gricean maxim of quantity appeared to lead participants away from the strongest possible set-theoretic translation for the particular formal statements corresponding to I and O premisses. Logically speaking, the meaning of a particular affirmative statement, " $\exists t : T \bullet A(t) \wedge B(t)$ ", does unquestionably entail the assertion "At least one A is a B", as endorsed by nearly three quarters of participants. However, it also entails the possibility that "All of the As might be Bs", which, providing participants had abided by the Gricean maxim of manner and made their interpretations as informative and unambiguous as was necessary for the purposes of the experiment, would have led them to the strongest possible set-theoretic interpretation, "At least one (possibly all) As are Bs", as endorsed by

roughly one fifth of participants. Similarly, the meaning of a particular negative, " $\exists t : T \bullet A(t) \wedge \neg B(t)$ ", does entail the assertion "At least one A is not a B", however, it also entails the possibility that "All of the As might not be Bs", as endorsed by roughly one fifth of participants. The high rate of subset translations of these formal expressions is perhaps attributable to people's strong inclination to draw partitive interpretations of the "some" and "some ... not" quantifiers in everyday communication, where, according to Gricean convention, the truth of one should imply the truth of the other (Newstead, 1989). The relatively high rates of precise, set-theoretic interpretations observed for the formal universal premisses, as compared with those for the particular premisses, is supported by a previous natural language based study (Neimark and Chapman, 1975) in which participants indicated their interpretations using Euler Circle combinations.

We postulate here whether participants' seemingly ambiguous understandings of the formal expressions with particular moods might be at least partly attributable to the way in which the existential quantifier is normally introduced in the textbooks on the Z notation. Most popular undergraduate texts (for example: Diller, 1994; Potter et al., 1996) state that the " $\exists$ " quantifier can be paraphrased as the English expression "there is" or "there exists" (which is easily memorisable from its "reverse E" symbology), but fail to mention that, unless explicitly predicated not to do so, the existentially quantified element of a set might refer to the entirety of the set that it represents. So, when an individual perceives the " $\exists$ " quantifier, he or she is inclined to assume a singular, rather than multiple, referent, "there exists one or several, but not all", which contrasts markedly with the " $\forall$ " quantifier, where "all", or "for every", is definitively asserted. But, when an individual perceives a statement beginning with " $\neg \exists$ ", he or she is inclined to assume the equally definitive "there does not exist", or "none", which may account for the high rate of unambiguous set-theoretic interpretations in the corresponding background task.

Based on the results of a previous study, Erickson (1974) proposes a three-part theoretic model of syllogistic reasoning which focuses on the ways in which people mentally represent given premiss pairs. Firstly, the "complete" model claims that an individual considers all possible representations of the combined premisses then selects a conclusion that encompasses all of the possible combinations. Under this model, errors become explainable as a consequence of the individual's inadequate interpretation of the represented models. Secondly, the "random" model claims that an individual considers only one combination of diagrams, with all possible combinations being equally likely, and gives a conclusion based on the one combination considered. Here, errors become explainable as a consequence of either inadequate interpretation or failure to consider all possible combinations. Thirdly, the "differential" model claims that the kind of combination considered will vary from person to person, and that some people will consider more than one of the possible combinations. Here, errors become explainable as a result of personal preferences, inadequate interpretation or failure to consider all hypothetical possibilities.

Here, we interpret the results of the present study in light of the theory that people represent and process premiss information in ways analogous to those in which Euler Circles or Venn Diagrams are employed in mathematics (Erickson, 1974; 1978). Inspection of those tasks in which at least three participants gave the same non-logical response seems to suggest two marked trends which could account for a large proportion of participants' errors. Firstly, where the syllogism is indeterminate, errors may be attributed to the participant's failure to represent all possible premiss combinations mentally and their endorsement of a conclusion which was consistent with their somewhat restricted interpretation. This trend seems most evident in participants' responses to the abstract syllogism OA1 and thematic syllogisms IE2 and IE4. Figure 19 shows how participants' apparent bias towards one representation of the thematic IE2 premiss pair prevented them from

finding a counter example to the erroneous interpretation endorsed by one quarter of participants, "No drunkards are scientists."

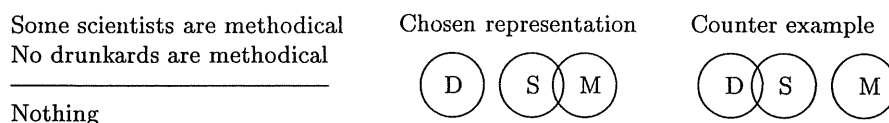


Figure 19: Set-theoretic representations of the IE2 syllogism

Secondly, where the syllogism is determinate, errors may be attributed to the participants' failure to adopt an appropriate A, E, I or O interpretation of correctly represented premiss combinations. This trend is most evident in responses to the thematic syllogisms AO2, EA1, EI3 and EI4 (unbelievable). In order to be certain that the determinate conclusion drawn from a given premiss pair is logically entailed under the set-theoretic approach, the reasoner must initially construct mental representations of all possible premiss combinations in order to determine whether any counter examples exist to putative conclusions. However, this systematic process of constructing premiss combinations and searching for counter examples can demand more mental effort than the reasoner is willing to expend (Barston, 1986; Johnson-Laird and Bara, 1984). Figure 20 illustrates how one quarter of participants appeared to respond based on probabilistic guesswork rather than construct all of the possible premiss representations which followed from the thematic EI3 premiss pair. Here, it is postulated that participants' failure to conduct an exhaustive analysis of this kind was responsible for one quarter of participants failing to derive the determinate interpretation, "Some conductors are not woods", which is consistent with all five representations.

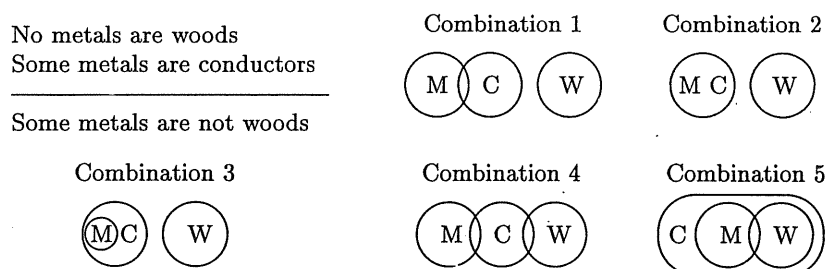


Figure 20: Possible set-theoretic representations of the EI3 premisses

Owing to the generality of its predictions, it is almost inconceivable that the erroneous responses observed in any syllogistic study could not be explained in terms of Erickson's three-part model. Errors in any deductive task can, after all, be ascribed to the possibility that the participant failed to consider the correct conclusion and it could be difficult to disprove this claim, even with the benefit of thorough post-hoc interrogation. Although the results of the present study are almost certainly explainable in terms of Erickson's model, they give rise to a much simpler, but equally general, account of erroneous performance in the syllogistic task: for indeterminate syllogisms, errors may be ascribed to the individual's failure to represent mentally all hypothetical possibilities, while for determinate syllogisms, errors may be ascribed to the individual's failure to consider appropriate A, E, I or O interpretations of represented possibilities. Overall, the results are generally consistent with the view that the difficulty of a syllogism increases along with the number of ways in which its premisses can be represented (Ceraso and Provitera, 1971; Erickson, 1974), and, notably, suggest that participants experienced most difficulty in drawing particular negative interpretations from represented premisses.

From our discussion of participants' performance in the present study, it should be clear that their responses were influenced by a variety of factors. Similarly, it is postulated here that participants' errors were not in general attributable to single, independent causes, but rather to the combination of several non-logical reasoning heuristics or biases. Furthermore, it is postulated that the factors which incited errors differed between participants. A summary of the factors which we believe to have contributed towards participants' endorsement of erroneous conclusions during the present study is given in Figure 21. It should be noted that many of these non-logical heuristics and biases are hypothesised in the cognitive literature to exert a central and dominating influence on human reasoning processes both during syllogistic studies and everyday communicative experience.

1. Failure to consider hypothetical possibilities
2. Failure to consider A, E, I or O interpretations of representations
3. Belief bias
4. Adherence or non-adherence to Gricean conventions
5. Propositional conclusion bias
6. Atmosphere effect or matching bias
7. The "Same M" fallacy
8. Probabilistic or caution bias
9. Illicit premiss conversion
10. Ambiguous set-theoretic or non-logical interpretations
11. Figural bias
12. Formalisation of the syllogistic task

Figure 21: Possible causes of error in the present study

Table 8 contains English translations of those abstract syllogisms in which three or more participants succumbed to the same non-logical response and gives a speculative list of possible causes for these errors, according to participants' responses. Inspection of the table suggests that most erroneous responses to the abstract tasks are explainable in terms of the cognitive theories of probabilistic bias, the "Same M" fallacy, and ambiguous set-theoretic interpretations of given premisses.

TABLE 8  
Abstract syllogisms which incited most non-logical responses

Task	Premises	Logical Response [Erroneous Response]	Possible Causes
AA4	All A are B, All B are C	Nothing [Some A are B]	4, 5, 7, 8, 12
EA3	No B are C, All B are A	Nothing [No A are C]	6, 7, 9, 10, 12
EA4	No A are B, All B are C	Nothing [Some C are not A]	4, 5, 7, 8, 10, 12
EI4	No A are B, Some B are C	Some A are not C [Nothing]	2, 8, 10
OA1	Some B are not C, All A are B	Nothing [Some A are not C]	4, 5, 6, 7, 8, 9, 10, 11
OA3	Some B are not A, All B are C	Some C are not A [Nothing]	2, 8, 9

Note: Numbers of possible causes relate to the list presented in Figure 21.

Table 9 delineates those thematic syllogisms in which three or more participants succumbed to the same non-logical response and presents a speculative list of possible causes for these errors, according to participants' responses. In common with participants' errors during the abstract tasks, the table reveals that participants' erroneous responses during the thematic tasks are explainable in terms of the cognitive theories of probabilistic bias, the "Same M" fallacy and ambiguous set-theoretic interpretations. But, in contrast to the observed errors for the abstract syllogisms, Table 9 suggests that belief bias, adherence or non-adherence to Gricean convention and propositional bias were much more influential on participants' performance during the thematic tasks. This trend appears consistent with the view that, in several cases, the meaningful nature of the syllogistic terms, when combined with prior beliefs relating to these terms and the natural tendency to respond according to conventions of everyday linguistic usage, which were possibly cued by the realistic nature of the thematic material, may have distorted the logical demands of the task.

TABLE 9  
Thematic syllogisms which incited most non-logical responses

Task	Premises	Logical Response [Erroneous Response]	Possible Causes
AA4	All bank managers are responsible, All responsible people are trustworthy	Nothing [Some trustworthy people are bank managers]	3, 4, 5, 7, 8, 12
AO2	All honest people are hard workers, Some politicians are hard workers	Some politicians are not honest [Nothing]	2, 7, 8, 10
IE2	Some scientists are methodical, No drunkards are methodical	Nothing [No drunkards are scientists]	1, 3, 4, 5, 6, 7, 8, 10
IE4	Some edible foods are vegetables, No vegetables are minerals	Nothing [No minerals are edible]	1, 4, 5, 6, 7, 8, 9, 10
EI3	No metals are woods, Some metals are conductors	Some conductors are not woods [Nothing]	2, 8
EI4*	No disloyal people are married, Some married people are traitors	Some traitors are not dis- loyal [Nothing]	2, 8
EA1	No millionaires are poor, All rich people are millionaires	No rich people are poor [Nothing]	2, 8
EA3	No churchgoers are atheists, All churchgoers are devout people	Nothing [No devout people are atheists]	3, 4, 5, 6, 7, 8, 10, 12
EA4	No oranges are apples, All apples are fruits	Nothing [Some fruits are not oranges]	3, 4, 5, 7, 8, 9, 10, 12

*Note:* Numbers of possible causes relate to the list presented in Figure 21. Premises leading to unbelievable conclusions are indicated with an asterisk.

## 5 Conclusions

"Aristotle, after all, invented the syllogisms as a means of enabling people to extract the logically necessary information from discourse and thus to loose themselves from the interpretive acquiescence that language invites. To the extent that people nevertheless perceive and treat the syllogisms as discourse, the system cannot serve its purpose. How could this problem be remedied? One possibility might be to make the system less seductively language-like. One might recast it on terms of, say, propositional logic. The major drawback to this solution is that the logic would remain relatively inaccessible except to reasoners with special training" (Adams, 1984, p.303-304).

It is postulated that one of the main causes of error in studies of syllogistic reasoning is the inability of participants to determine whether to apply the laws of natural language or the laws of logic (Politzer, 1986). Given participants' logical training and the explicitly logical nature of the tasks in the present study, one would have expected participants to be cued into using the laws of logic throughout. The fact that participants' levels of correctness were generally more pronounced than has been observed for novice reasoners in previous natural language based studies suggests that the laws of logic did exert a dominating influence on reasoning. However, the high rates of error observed for certain tasks, which are consistent with cognitive theories of erroneous reasoning in natural language contexts, suggests that computing scientists, despite their logical training and the clearly logical nature of the tasks, are prone to disregard logic in favour of non-logical biases and heuristics, including those based on the pragmatic conventions governing everyday communication. Thus, with respect to Adams' (1984) hypothesis, formalisation in terms of a notation with strong foundations in propositional logic appears to provide only a partial remedy for people's errors in the syllogistic task; it does not seem capable of preventing participants completely, even those with the "special training", from applying inappropriate language conventions from ordinary discourse.

Given a problem which calls for deductive reasoning, such as a syllogism, both logical and non-logical tendencies compete for control of our cognitive processes, but it is only by following logical principles that we can be certain of arriving at the correct solution (Evans et al., 1983; Politzer, 1986). Perhaps because of their inability to differentiate between the laws of language and the laws of logic, participants in the present study seemed strongly inclined to adhere to Gricean maxims of everyday communication. This trend was particularly noticeable under the thematic condition where the presentation of everyday terms seemed almost to cue the conventions of everyday linguistic usage and lead them away from the logic of the tasks. Many of the observed errors appear attributable to participants' failure to recognise that the Gricean maxims of quantity and relation, in particular, were incompatible with the logical forms of analysis required by the tasks. It was perhaps participants' adherence to the maxim of relevance, for example, which incited numerous determinate conclusions in response to indeterminate tasks. However, participants' conformance with Gricean convention was not necessarily detrimental to their performance on all occasions. Adherence to the maxim of quantity during the background tasks, for example, led most participants to unambiguous set-theoretic interpretations of formal statements with A and E moods, and, during the main tasks, led to most participants endorsing strong conclusions where weaker ones were also available.

In actuality, then, the predisposition to conform with Gricean convention may have led participants away from the logic of some tasks, but fortuitously pointed them towards the correct conclusion in others. Participants' errors in many cases, then, appear to have stemmed from their belief that pragmatic conventions of everyday communication were universally applicable. Previous studies suggest that it is only through environmental and educational experience that people learn to appreciate fully the logical meanings of the predicate quantifiers (Politzer, 1990) and logical connectives (Neimark and Chapman, 1975). It is only through experience, then, that the individual can learn the differences between the laws of logic and the laws of language. Moreover, it is only through experience, that the individual can appreciate the difference between those occasions in which a purely logical or an informal pragmatic performance is appropriate. The development of such an appreciation appears to be a necessary pre-requisite for the syllogistic task. Such findings are consistent with previous studies of bilingualism. Dalrymple-Alford (1968), for example, argues that the individual has only a single psycholinguistic system in which knowledge of both his or her native and secondary languages reside, and that neither language can be completely blocked or disabled in situations where both



can be applied. Whilst Kiyak (1982) suggests that it is only through increased language familiarity that the individual can learn to ignore cues which evoke the inappropriate language and begin to communicate without interlingual interference.

In general, erroneous reasoning leads to erroneous decisions, and the entire history of software development has shown us that incorrect development decisions can lead to the introduction of defects in delivered systems (Potter et al., 1996; Sommerville, 1992). Software engineering has always been a human oriented activity and it is likely to remain so, at least for the foreseeable future. Thus, the potential for human error in the software development process is likely to remain despite the use of formal methods. Although it is argued that formalisation increases overall confidence in the integrity of delivered systems (Ince, 1992; Liskov and Berzins, 1986), the results of the present study suggest that, although the users of formal methods are mostly logical in their reasoning about quantified formal expressions, they are still occasionally liable to err in ways similar to those observed for natural language based notations. That the results point to the possible existence of non-logical encoding, processing and response biases suggests that the psychological causes of error in formal syllogistic reasoning are indeed deep-rooted. If the same can be said of human reasoning in formal contexts generally, as seems to be suggested by the results, then it is likely that formal specification will always remain, to some extent, vulnerable to the fallibility of human reasoning, and "if it is impossible to guarantee the elimination of errors, then we must discover more effective ways of mitigating their consequences in unforgiving situations" (Reason, 1990, p.148). It is disconcerting to think that software developers will exhibit similar, or even increased, potentials for error in critical industrial projects, where the correct decisions are rarely offered explicitly in the form of multiple-choice answers, and where the ramifications of erroneous reasoning are much more serious than in laboratory based studies. It is argued in the software community that the primary purpose of a software specification document should be as a medium for communication (Barroca and McDermid, 1992; Imperato, 1991). We agree that formal software specifications should, in general, be expressed in ways which are clearly comprehensible to all members of their intended audiences, but in addition we propose that they have minimal potential for inciting erroneous development decisions, because it is only when one appreciates the repercussions that erroneous development decisions can have on software projects and the quality of their delivered products that one can begin to understand the need for capturing and verifying the reasoning processes of software developers.

The increasing interest in formal methods being shown by the software community (Bowen and Hinchey, 1994; Oakley, 1990) may be partly attributable to the widespread belief that it is easier to reason about formal software specifications than conventional natural language based specifications (Thomas, 1995). However, the software community has been slow to support this, and other claims pertaining to the use of formal methods, with empirical evidence (Craigien et al., 1995; Fenton, 1996). The research methodology followed in the present study, however, demonstrates an approach via which claims such as these can be subjected to independent and objective examination. Rather than being based on subjective personal belief, which might not accurately reflect objective reality, the lines of inquiry pursued in the present research stem from the well supported findings of previous cognitive studies. Moreover, rather than using isolated case studies, from which it can be difficult to extrapolate results, the methodology borrows standard experimental procedures from the domain of cognitive science. This approach is advantageous over conventional software research methodologies in that it generates theories which are "grounded" in the observed data (Glaser and Strauss, 1968), which can subsequently be used to refine initial hypotheses and generate new theories in a fashion analogous to the Popperian "underlying pattern of continuous development" (Magee, 1985).

The finalised theories and data that emerge from such a line of investigation, then, provide an empirical basis via which the psychological claims of the software community can be subjected to empirical scrutiny.

The overall aim of the present research, of which the study reported here is a part, is to identify combinations of grammatical constructs which are particularly susceptible to incite errors and biases when people are reasoning with formal specifications. As compensatory measures are then introduced, this can help to reduce the potential for human error in the software development process. After all, if we know when and where errors are most likely to occur then erroneous development decisions can be pre-empted and the numbers of defects introduced into "finished" software systems reduced. Previously, we have borrowed from cognitive science the relevant theoretical knowledge, experimental methodology and statistical procedures in order to determine the precise conditions under which trained users are particularly susceptible to error and bias when reasoning about formal Z specifications containing conditionals (Loomes and Vinter, 1997; Vinter et al., 1997a), disjunctives and conjunctives (Vinter et al., 1997b). During the present study, we have extended our line of inquiry to include the existential and universal quantifiers. In conducting these studies, we have demonstrated the feasibility of a cognitive approach to evaluating the formal specification process which, we are convinced, is at least as important as the results themselves.

## References

- Adams, M.J. (1984). Aristotle's logic. In G.H. Bower (Ed.), *The Psychology of Learning and Motivation*, 18, London: Academic Press.
- Barden, R., Stepney, S., Cooper, D. (1992). The Use of Z. In J.E. Nicholls (Ed.), *Proceedings of the Sixth Annual Z User Meeting, York 1991*, 99-124, London: Springer-Verlag.
- Barroca, L.M. and McDerimid, J.A. (1992). Formal methods: Use and relevance for the development of safety-critical systems. *The Computer Journal*, 35, 6, 579-599.
- Barston, J.L. (1986). *An investigation into belief biases in reasoning*. Unpublished PhD thesis, University of Plymouth.
- Begg, I. and Denny, J. (1969). Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning errors. *Journal of Experimental Psychology*, 81, 351-354.
- Begg, I. and Harris, G. (1982). On the interpretation of syllogisms. *Journal of Verbal Learning and Verbal Behaviour*, 21, 595-620.
- Beth, E.W. and Piaget, J. (1966). *Mathematical Epistemology and Psychology*. Dordrecht: Reidel.
- Bowen, J.P. (1988). Formal specification in Z as a design and documentation tool. *Second IEE/BCS Conference, Software Engineering 88*, 164-168.
- Bowen, J.P. and Hinchey, M.G. (1994). Seven more myths of formal methods: Dispelling industrial prejudice. In T. Denvir, M. Naftalin and M. Bertran (Eds.), *FME'94: Industrial Benefit of Formal Methods*, 105-117, LNCS 873, London: Springer-Verlag.
- Bowen, J.P. and Stavridou, V. (1993). Safety-critical systems, formal methods and standards. *Software Engineering Journal*, July 1993.
- 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.
- Chapman, L.J. and Chapman, J.F. (1959). Atmosphere effect re-examined. *Journal of Experimental Psychology*, 58, 220-226.
- Ceraso, J. and Provitera, A. (1971). Sources of error in syllogistic reasoning. *Cognitive Psychology*, 2, 400-410.

- Cohen, B. (1989). Justification of formal methods for system specification. *Software Engineering Journal*, 1, 26-35, January 1989.
- Craigien, D., Gerhart, S. and Ralston, T. (1995). Formal methods technology: Impediments and innovation. In M.G. Hinchey and J.P. Bowen (Eds.), *Applications of Formal Methods*, 399-419, Hemel Hempstead: Prentice-Hall.
- Dalrymple-Alford, E.C. (1968). Interlingual interference in a color-naming task. *Psychonomic Science*, 10, 215-216.
- Dean, N. and Hinchey, M.G. (1996). Formal methods and modeling in context. In C.N. Dean and M.G. Hinchey (Eds.), *Teaching and Learning Formal Methods*, 99-116, London: Academic Press.
- Dickstein, L.S. (1975). Effects of instructions and premiss order on errors in syllogistic reasoning. *Journal of Experimental Psychology: Human Learning and Memory*, 104, 376-384.
- Dickstein, L.S. (1976). Differential difficulty of categorical syllogisms. *Bulletin of the Psychonomic Society*, 8, 330-332.
- 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.
- Dickstein, L.S. (1981). The meaning of conversion in syllogistic reasoning. *Bulletin of the Psychonomic Society*, 18, 3, 135-138.
- Diller, A. (1994). *Z. An Introduction to Formal Methods*. Second edition, Chichester: Wiley and Sons.
- Dominowski, R.L. (1995). Content effects in Wason's selection task. In S.E. Newstead and J.St.B. Evans (Eds.), *Perspectives on Thinking and Reasoning. Essays in Honour of Peter Wason*. Hove: Erlbaum.
- Empson, W. (1965). *Seven Types of Ambiguity*. Harmondsworth: Penguin.
- 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: Erlbaum.
- 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. (1972). Interpretation and matching bias in a reasoning task. *Quarterly Journal of Experimental Psychology*, 24, 193-199.
- 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.
- Fenton, N. (1996). The empirical basis for software engineering. In A. Melton (Ed.), *Software Measurement*, 197-217, London: Thomson.
- Fenton, N.E. and Pfleeger, S.L. (1996). *Software Metrics. A Practical and Rigorous Approach*. London: ITC Press.
- Gehani, N. (1986). Specifications: Formal and informal - a case study. In N. Gehani and A.D. McGettrick, *Software Specification Techniques*, Wokingham: Addison-Wesley.
- 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.
- Glaser, B.G. and Strauss, A.L. (1968). *The Discovery of Grounded Theory. Strategies for Qualitative Research*. London: Weidenfeld and Nicolson.

- Grice, H.P. (1975). Logic and conversation. In P. Cole and J.L. Morgan (Eds.), *Syntax and Semantics, 3: Speech Acts*, 41-58, New York: Academic Press.
- 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. and Michael, M. (1956). The influence of attitudes on syllogistic reasoning. *Journal of Social Psychology*, 44, 115-127.
- Imperato, M. (1991). *An Introduction to Z*. Bromley: Chartwell-Bratt.
- Ince, D.C. (1992). *An Introduction to Discrete Mathematics, Formal Specification and Z*. Second edition, 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.
- James, W. (1950). *Principles of Psychology, Volume 2*. London: Dover Publications.
- Janis, I. 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. (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 Bara, B. (1984). Syllogistic inference. *Cognition*, 16, 1-61.
- Johnson-Laird, P.N. and Steedman, M.J. (1978). The psychology of syllogisms. *Cognitive Psychology*, 10, 64-99.
- 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., Byrne, R.M.J. and Tabossi, P. (1989). Reasoning by model: The case of multiple quantification. *Psychological Review*, 96, 658-673.
- Johnstone, P.T. (1987). *Notes on Logic and Set Theory*, Cambridge: Cambridge University Press.
- Kiyak, H.A. (1982). Interlingual interference in naming colour words. *Journal of Cross-Cultural Psychology*, 13, 1, 125-135.
- 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. (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.
- Lukasiewicz, J. (1967). *Aristotle's Syllogistic. From the Standpoint of Modern Formal Logic*. Second Edition, Oxford: Clarendon Press.
- MacKensie, D. (1992). Computers, formal proof, and the law courts. *Notices of the American Mathematical Society*, 39, 9, 1066-1069, November 1992.
- Magge, B. (1985). *Popper*, London: Fontana Press.
- Meyer, B. (1985). On formalism in specifications. *IEEE Software*, 2, 1, 6-26.
- 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.
- Moxey, L.M. and Sanford, A.J. (1993). *Communicating Quantities: A Psychological Perspective*. Hove: Erlbaum.

- Neimark, E.D. and Chapman, R.H. (1975). Development of the comprehension of logical quantifiers. In R.J. Falmagne (Ed.), *Reasoning: Representation and Process in Children and Adults*, 135-151, Hillsdale: Erlbaum.
- Newstead, S.E. (1989). Interpretational errors in syllogistic reasoning. *Journal of Memory and Language*, 28, 78-91.
- Newstead, S.E. (1990). Conversion in syllogistic reasoning. In K.J. Gilhooly, M.T.G. Keane, R.H. Logie, and G. Erdos, *Lines of Thinking: Reflections on the Psychology of Thought. Volume 1. Representation, Reasoning, Analogy and Decision Making*, 73-84, Chichester: Wiley.
- Newstead, S.E. (1995). Gricean implicatures and syllogistic reasoning. *Journal of Memory and Language*, 34, 5, 644-664.
- Newstead, S.E. and Griggs, R.A. (1983). Drawing inferences from quantified statements: A study of the square of opposition. *Journal of Verbal Learning and Verbal Behaviour*, 22, 535-543.
- Newstead, S.E., Griggs, R.A. and Chrostowski, J.J. (1984). Reasoning with realistic disjunctives. *Quarterly Journal of Experimental Psychology*, 36A, 611-627.
- Norcliffe, A. and Slater, G. (1991). *Mathematics of Software Construction*. London: Ellis Harwood.
- 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, *Lines of Thinking: Reflections on the Psychology of Thought. Volume 1. Representation, Reasoning, Analogy and Decision Making*, 125-138, Chichester: Wiley.
- Oakley, B. (1990). Opening address: The state of use of formal methods. In J.E. Nicholls (Ed.), *Proceedings of the Fourth Annual Z User Meeting, Oxford 1989*, 1-5, London: Springer-Verlag.
- Politzer, G. (1986). Laws of language use and formal logic. *Journal of Psycholinguistic Research*, 15, 47-92.
- Politzer, G. (1990). Immediate deduction between quantified sentences. 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*, 85-98, Chichester: Wiley.
- Potter, B., Sinclair, J. and Till, D. (1996). *An Introduction to Formal Specification and Z*. Second edition, Hemel Hempstead: Prentice-Hall.
- Reason, J. (1990). *Human Error*. Cambridge: Cambridge University Press.
- Reichenbach, H. (1966). *Elements of Symbolic Logic*. London: Collier-Macmillan.
- Revlín, R. and Leirer, V.O. (1980). Understanding quantified categorical expressions. *Memory and Cognition*, 8, 447-458.
- Revlín, R., Leirer, V., Yopp, H. and Yopp, R. (1980). The belief-bias effect in syllogistic reasoning: The influence of knowledge on logic. *Memory and Cognition*, 8, 584-592.
- Revlis, R. (1975a). Two models of syllogistic inference: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behaviour*, 14, 180-195.
- Revlis, R. (1975b). Syllogistic reasoning: Logical decisions from a complex data base. In R.J. Falmagne (Ed.), *Reasoning: Representation and Process in Children and Adults*, 93-133, Hillsdale: Erlbaum.
- Roberge, J. (1970). A re-examination of the interpretation of errors in formal syllogistic reasoning. *Psychonomic Science*, 19, 331-333.
- Roberge, J.J. (1976). Effects of negation on adults' disjunctive reasoning abilities. *Journal of General Psychology*, 24, 87-91.
- Roberge, J.J. (1978). Linguistic and psychometric factors in propositional reasoning. *Quarterly Journal of Experimental Psychology*, 30, 705-716.
- Ross, W.D. (Ed.). (1949). Aristotle, *Analytica priora et posteriora*. Oxford: Clarendon Press.

- Russell, B. (1994). *History of Western Philosophy and its Connection with Political and Social Circumstances from the Earliest Times to the Present Day*. London: Routledge.
- Sells, S.B. (1936). The atmosphere effect: An experimental study of reasoning. *Archives of Psychology*, 200, 1-72.
- Sells, S.G. and Koob, H.F.A. (1937). A classroom demonstration of "atmosphere effect" in reasoning. *Journal of Educational Psychology*, 28, 514-518.
- Sheppard, M. and Ince, D. (1993). *Derivation and Validation of Software Metrics*. Oxford: Oxford University Press.
- Simpson, M. and Johnson, D. (1966). Atmosphere and conversion errors in syllogistic reasoning. *Journal of Experimental Psychology*, 72, 197-200.
- Smith, N.K. (Ed. and Trans., 1993). I. Kant, *Immanuel Kant's Critique of Pure Reason*. London: Macmillan.
- Sommerville, I. (1992). *Software Engineering*. Fourth Edition, Wokingham: Addison-Wesley.
- Spivey, J.M. (1992). *The Z Notation: A Reference Manual*. Second Edition, Prentice-Hall.
- Strawson, P.F. (1966). *Introduction to Logical Theory*. London: Methuen.
- Thistlethwaite, D. (1950). Attitude and structure as factors in the distortion of reasoning. *Journal of Abnormal and Social Psychology*, 45, 442-458.
- Thomas, M. (1995). Formal methods and their role in developing safe systems. *High Integrity Systems Journal*, 1, 5, 447-451.
- Traub, B.H. (1977). A set theory approach to deduction with meaningful syllogisms. Unpublished PhD thesis, Ohio State University.
- Tsal, Y. (1977). Symmetry and transitivity assumptions about a nonspecified logical relation. *Quarterly Journal of Experimental Psychology*, 29, 677-684.
- Turner, G.W. (1986). *Stylistics*. Harmondsworth: Penguin.
- 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. (1997a). *Conditional Reasoning in Language and Logic: Transfer of Non-logical Heuristics?* University of Hertfordshire, Division of Computer Science, Technical Report No. 276.
- Vinter, R.J., Loomes, M.J. and Kornbrot, D.E. (1997b). *A Study of Disjunctive and Conjunctive Reasoning in Formal Logic*. University of Hertfordshire, Division of Computer Science, Technical Report No. 298.
- Wason, P.C. and Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, 23, 63-71.
- Wetherick, N.E. (1993). Psychology and syllogistic reasoning. *Philosophical Psychology*, 6, 4, 423-440.
- Wetherick, N.E. and Gilhooly, K.J. (1990). Syllogistic reasoning: effects of premise order. In K.J. Gilhooly, M.T.G. Keane, R.H. Logie, and G. Erds (Eds.), *Lines of Thinking: Reflections on the Psychology of Thought. Volume 1. Representation, Reasoning, Analogy and Decision Making*, 99-108, Chichester: Wiley.
- Wilkins, M.C. (1928). The effect of changed material on ability to do formal syllogistic reasoning. *Archives of Psychology*, 16, 102, 1-83.
- Woodworth, R.S. and Sells, S.B. (1935). An atmosphere effect in syllogistic reasoning. *Journal of Experimental Psychology*, 18, 451-460.