John Benjamins Publishing Company

This is a contribution from *Pragmatics & Cognition 11:2*
© 2003. John Benjamins Publishing Company

This electronic file may not be altered in any way.
The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.
Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute.
For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).
Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

Mental models theory and relevance theory in quantificational reasoning*

Steve Nicolle
SIL International, Kenya

Human reasoning involving quantified statements is one area in which findings from cognitive psychology and linguistic pragmatics complement each other. I will show how mental models theory provides a promising account of the mechanisms underlying peoples' performance in three types of reasoning tasks involving quantified premises and conclusions. I will further suggest that relevance theory can help to explain the way in which mental models are employed in the reasoning processes. Conversely, mental models theory suggests that human reasoning typically does not involve deductive rules, which in turn entails a modification to the nature of the deductive processes proposed by relevance theory. The mechanism proposed by mental models theory also helps to clarify the nature of the relevance theory distinction between conceptual and procedural information.

1. Introduction

In this paper I will discuss experimental evidence concerning how people reason when presented with arguments involving quantified premises and conclusions. I will look specifically at three kinds of reasoning tasks: reasoning about possibility and necessity, reasoning with believable and unbelievable conclusions, and reasoning with modal conclusions. For each type of task I will argue that the experimental results are what one would expect if the reasoning process is based upon the manipulation of mental models and governed by the principle of relevance. That is, I will argue that mental models theory and relevance theory together go a long way towards providing a satisfactory explanation of the results obtained in these three distinct types of quantificational reasoning task.

At the same time, analysis of data from these reasoning tasks necessitates certain modifications to both relevance theory and mental models theory.

Pragmatics & Cognition 11:2 (2003), **345–378.**
ISSN 0929–0907 / E-ISSN 1569–9943 © John Benjamins Publishing Company

© 2003. John Benjamins Publishing Company
All rights reserved

Underlying the analyses reported here is the assumption, shared by relevance theory, mental models theory and the related verbal reasoning theory of Polk and Newell, that "the most important cognitive processes in deduction are the same ones that are used in language comprehension and generation" (Polk and Newell 1995: 535). Because this is an assumption, and therefore subject to potential falsification at any time, the conclusions drawn in this study can never be more than suggestive. This said, if similar cognitive processes are in fact at work during language comprehension and deduction, then the success of mental models theory in explaining the patterns of errors found in these reasoning tasks casts doubt on the relevance theory claim (Sperber and Wilson 1986: 85) that inferential processing proceeds on the basis of deductive rules such as *modus ponendo ponens* and *modus tollendo ponens* (see also Dascal 1987 for a discussion of the role of language in reasoning as an alternative to rule-based approaches).

To say that the cognitive processes used in deduction are the same as those used in utterance interpretation is one thing; it is another thing to say that the way such processes are applied in logical reasoning tasks is the same as the way they are applied in utterance interpretation. According to Johnson-Laird, the ability to construct and manipulate mental models underlies both utterance interpretation (and production) and logical reasoning; however, utterance interpretation requires only a single mental model to be entertained at any one time (which is why it can be so rapid, according to Johnson-Laird, 1983: 128), whereas logical reasoning requires "a knowledge of the semantic principle of validity and the ability to put it into practice in the search for counter examples" (Johnson-Laird 1986: 37). Detailed analysis in terms of mental models and relevance theory of the patterns of errors found in people's performance in reasoning tasks suggests that the way cognitive processes are applied in such tasks is consistent with the way such processes are applied in utterance interpretation rather than in successful (valid) logical reasoning.[1]

A detailed introduction to mental models theory and relevance theory is beyond the scope of this paper, so in Sections 1.1 and 1.2 I will merely highlight the components of each that have the greatest bearing on this study.

1.1 Relevance theory

The central claim of relevance theory[2] (Sperber and Wilson 1986) is that human information processing, of which utterance interpretation is a particular case, adheres to a cost versus benefit principle according to which people, as

© 2003. John Benjamins Publishing Company
All rights reserved

information processors, seek adequate contextual effects (that is, modifications to an individual's information base) for minimum processing effort. This is referred to as 'optimal relevance'. Utterance interpretation differs from other cases of reasoning in that an utterance is an instance of an ostensive stimulus, that is, it is a stimulus intentionally created not only to communicate but also to be recognised by the intended audience as having been created with the intention to communicate. Every ostensive stimulus is presumed to come with a presumption of optimal relevance: it communicates (to its intended recipients) the assumption that it will yield adequate contextual effects in a way that minimises processing effort.

An addressee acting in accordance with the principle of relevance should bring the inferential process to an end once adequate contextual effects have been achieved, thereby minimising processing effort. Conversely, if an ostensively communicated stimulus is structured in such a way as to require additional processing effort, an addressee processing that utterance in accordance with the principle of relevance should look for additional contextual effects in order to justify the additional processing effort.

1.2 Mental models theory

Whilst relevance theory aims to describe the *principles* governing human cognition, mental models theory (Johnson-Laird 1983, 1986; Johnson-Laird and Bara 1984; Johnson-Laird and Byrne 1991; Johnson-Laird, Byrne, and Schaeken 1992) aims to describe the *mechanisms* underlying human cognition. According to mental models theory, human reasoning involves the construction of propositional representations of situations that function as premises in an argument. These propositional representations are mapped into mental models of the world with respect to which the representations are to be interpreted; that is, "propositional representations are interpreted with respect to mental models" (Johnson-Laird 1983: 156). Conclusions are then drawn on the basis of consistency with these mental models.

Mental models theory proposes that successful deductive reasoning requires three stages:

i. A model of the premise(s) is constructed.
ii. An informative (non-repetitious) conclusion is drawn from this model.
iii. An alternative model is constructed in an attempt to falsify the initial conclusion.

© 2003. John Benjamins Publishing Company
All rights reserved

Reasoning through the performance of the first two stages only may lead to the generation of a valid conclusion, but does not guarantee it. Such 'short-circuited' reasoning does have the advantage, however, of requiring less effort than the performance of all three stages. During utterance interpretation, where conclusions must be drawn rapidly and where information is presented with the presumption that minimal processing effort is required (so that reasoners feel justified in accepting a conclusion based on the first model constructed), there are obvious benefits to 'short-circuited' reasoning. Consistently successful logical reasoning, on the other hand, requires that people search for counter examples to determine whether or not the initial conclusion can be falsified. The fact that performance on certain reasoning tasks is typically so poor, however, suggests that people typically do not attempt to falsify their initial conclusions. The experimental evidence presented in this paper suggests that whilst people do employ mental models in the reasoning process, they typically do not search for counter-example models in order to test an initial conclusion based on the first model constructed. That is, the heuristic usually applied is the one that is appropriate to utterance interpretation rather than the one that is guaranteed to lead to successful logical reasoning.[3]

1.3 Overview

Evans, Handley, Harper, and Johnson-Laird (1999) note that a conclusion can be accepted as possible but not necessary on the basis of stages (i) and (ii) alone, and their experiments demonstrate that people are more willing to endorse conclusions as possible than as necessary. This finding is consistent with the relevance theory claim that reasoning is guided by the search for optimal relevance. As Evans *et al.* note, much of our everyday reasoning, from choosing courses in a modular degree program to criminal law, is concerned with what is possible rather than with what is necessary. Endorsing a conclusion as possible rather than as necessary conforms to the relevance theory heuristic, since it minimises processing effort (only the first two stages of deductive reasoning are required) and will usually achieve adequate effects. Section 3 discusses Evans *et al.*'s results and a relevance theory interpretation of them in more detail.

However, as Evans *et al.* also observe, people can search for alternative models, even though they may require some additional motivation to do so. This may happen when a presented conclusion is unbelievable, a phenomenon known as 'belief bias'. It appears that in reasoning tasks where (under neutral

© 2003. John Benjamins Publishing Company
All rights reserved

conditions) people typically accept invalid arguments, error rates can be reduced if the conclusions are presented in the form of unbelievable statements. In such cases, people typically do search for counter-example models in order to falsify the presented conclusion. Two model-theoretic accounts of belief bias are compared in Section 4.

People are also more likely to search for alternative models when experimental instructions emphasise logical necessity to a sufficient degree (see also Newstead, Pollard, Evans, and Allen 1992: 276–281) or when conclusions are presented in the form of modal statements in which necessity corresponds to the use of the modal auxiliary *must*.[4] In Section 5, I argue that this is not unexpected given a relevance theory treatment of the semantics of modal auxiliaries and the notion of optimal relevance. Before looking at these different types of reasoning task, I will briefly introduce the study of arguments involving quantified premises and conclusions.

2. Syllogisms and immediate inference tasks

There are two main kinds of arguments involving quantified premises and conclusions which have been investigated by cognitive psychologists: syllogisms and immediate inference tasks. Syllogisms consist of two premises and a conclusion, each containing one of the quantifiers *All*, *No*, *Some* and *Some… not* with scope over 'terms'. Whilst syllogisms are deductions based on two premises, immediate inference tasks require a conclusion to be drawn from a single quantified premise (drawn from the same set of quantifiers). An important characteristic of syllogisms is that the quantifier *Some* is defined as meaning 'at least one and possibly all'. (The quantifiers are commonly referred to by the abbreviations A, E, I and O respectively, and the terms by lower case letter, so *All Cs are As* becomes *Aca*, *No Cs are As* becomes *Eca*, and so forth). One of the terms, traditionally referred to as the 'middle term', is common to both premises. The subject's task is to draw a conclusion linking the other two terms (called the 'end terms'). Traditionally, conclusions are drawn linking the end term in the second premise to the end term in the first premise; for example:

All Bs are As
Some Bs are Cs
Therefore, some Cs are As

© 2003. John Benjamins Publishing Company
All rights reserved

From their origins with Aristotle until the development of modern predicate calculi by Frege and by Russell and Whitehead, syllogisms were central to the study of logic (Lukasiewicz 1957 provides an account of the logical basis of syllogisms). More recently, syllogisms have attracted the attention of psychologists because of variations in people's performance in different syllogisms. Some syllogisms, such as the one illustrated above, have a high success rate, whilst other syllogisms are far harder for people to evaluate correctly. The factors with which difficulty and facility in syllogisms have been linked include the order of the terms in the premises and in the conclusions (Dickstein 1978; Johnson-Laird and Bara 1984; Johnson-Laird and Byrne 1991; Wetherick and Gilhooly 1990), elaboration of the premises (Ceraso and Provitera 1971; Newstead and Griggs 1999), the interpretation of *All* (Newstead 1995) and *Some* (Begg and Harris 1982; Begg 1987), which conclusions are presented (Hardman and Payne 1995), the believability of the conclusions (Newstead, Pollard, Evans, and Allen 1992; Rips 1994 and references therein), and subjects' own preferred reasoning strategies (Ford 1995).

Not surprisingly this substantial body of research has given rise to a number of different theories of syllogistic reasoning, most of which fall into one of three main classes (Newstead and Griggs 1999): mental logic (Rips 1994), which proposes that syllogistic reasoning proceeds through general purpose inference rules; response bias theories (e.g., Woodworth and Sells 1935; Wetherick and Gilhooly 1990), which propose that syllogisms are solved on the basis of prior beliefs rather than logic; and mental models theory, in which conclusions are drawn on the basis of models of the premises (see also Polk and Newell 1995 for a related approach). Of these theories, mental models theory has probably the widest range of applications (cf. Johnson-Laird 1995), is potentially falsifiable, and (with modifications) has arguably the least evidence against it (Klauer and Oberauer 1995; Newstead 1995; Girotto, Mazzocco, and Tasso 1997).

According to mental models theory, the quantified statements that constitute the premises in syllogisms may sometimes be represented by more than one mental model. For example, the universal quantifier *All* in 'All As are Bs' gives rise to two simple mental models[5] corresponding to the identity and subset relation respectively. The square brackets around a term indicate that this term is exhaustively represented (the notion of a simple mental model, and its application to premises containing the quantifier *Some* is discussed further in Section 5):

© 2003. John Benjamins Publishing Company
All rights reserved

All As are Bs	*(i) identity relation)*	*(ii) subset relation*
	[a] = [b]	[a] = b
	[a] = [b]	[a] = b
	[a] = [b]	b

A number of experimental studies on immediate inference tasks, in which conclusions are drawn from a single quantified premise, have been conducted. For example, given 'All As are Bs' as a premise, 'Some As are Bs' and 'Some Bs are As' are valid conclusions (trivially, so is 'All As are Bs') but 'All Bs are As' is not a valid conclusion since it is only supported by the identity model (i) but not by the subset model (ii).

In representing a syllogism, at least two simple mental models must be constructed, one for each premise. These then have to be combined and a conclusion drawn from the resulting combined model. In some syllogisms, a correct conclusion can be drawn from any one combined model; these are known as single model syllogisms. Other syllogisms, known as multiple model syllogisms, require consideration of more than one combined model in order to draw a valid conclusion. Errors can potentially arise at three places in the inferential process: (i) when subjects do not exhaustively represent the models of the premises (for example, by constructing only the identity model as a representation of 'All As are Bs'); (ii) when models are not combined properly; and (iii) when invalid conclusions are drawn from the resulting combined model. The research presented here points to the first potential source of error, not considering a sufficient number of models, as being responsible for many of the recorded errors.

To conclude this section, let us look at one sample syllogism.[6]

All As are Bs
All Bs are Cs

If only the identity relation is represented, the following combined model will result:

[a] = [b] = [c]
[a] = [b] = [c]
[a] = [b] = [c]

From this model, the conclusions 'All As are Cs' and 'All Cs are As' can be drawn. Only the former is valid however. 'All Cs are As' is the result of a 'conversion error' which results from only considering the identity model, and can be shown to be invalid by considering the subset model:

© 2003. John Benjamins Publishing Company
All rights reserved

[a] = [b] = c
[a] = [b] = c
[b] = c
[b] = c

By considering both combined models, we find that whilst 'All As are Cs' is a valid conclusion (that is, it is compatible with both models), 'All Cs are As' is not valid. However, another conclusion can be drawn linking C to A: 'Some Cs are As'. This is valid since it is compatible with both combined models given that *Some* is defined as 'At least one and possibly all'.

3. Possibility and necessity

Most studies of syllogistic reasoning have investigated people's performance in determining which conclusions are necessary given certain premises. Very few studies have investigated people's ability to determine whether conclusions are possible. This is curious, given that much of our everyday reasoning concerns the determination of what is possible rather than what is necessary. Facts which remain true irrespective of changing circumstances are typically given to us, and we are usually told if a course of action must or cannot be taken; we are rarely called upon to work out for ourselves whether such facts or actions follow as necessary conclusions from other propositions. Between the extremes of what must and cannot be the case, there is often a range of possibilities, and it is rare for people to be presented with all possible courses of action or interpretations of events.

For example, in choosing which courses to study and in which order as part of a university degree curriculum, a student will be constrained to a certain extent whilst also being allowed a certain amount of freedom in his or her choice. The regulations may stipulate that a particular number of courses must be taken at a given level, that at least a certain number must be within the student's main area of specialisation, and that some higher level courses require particular lower level courses to have been previously taken. There may also be certain courses that cannot be taken in the same term or year because of restrictions in the timetable. If the university authorities are communicating relevantly, all of this information will be given rather than left for the student to work out. Given these stipulations, it is up to each student to work out which desired combinations of courses are possible.

© 2003. John Benjamins Publishing Company
All rights reserved

Possibilities are thus clearly important in everyday reasoning, and unlike necessities they are generally computed rather than given. In relevance theoretic terms, determining whether or not a state of affairs is possible is likely to generate significant contextual effects. Determining necessity, on the other hand, is not the kind of computation that is performed often; in general, if it is necessary for an addressee to know whether a state of affairs is necessary rather than merely possible, a rational communicator will make this explicit rather than leaving it to the addressee to work it out. In this section, we will look, from a relevance theoretic perspective, at two studies which compare reasoning about necessity and possibility: Evans, Handley, Harper and Johnson-Laird (1999) and Evans, Handley and Harper (In press).

Evans *et al.* (1999: 1497) noted that according to mental models theory the third of the three stages of reasoning (constructing an alternative model in order to attempt to falsify an initial conclusion) is only necessary when people are asked to determine whether a conclusion is necessary. A conclusion can be possible if only one model supports it. From this, Evans *et al.* (1999: 1497) hypothesised that "[p]eople will be more willing to endorse conclusions as possible than as necessary". This hypothesis was supported for 'necessary problems' (in which the conclusion is true in all models of the premises; i.e., necessarily true), 'possible problems' (in which the conclusion is true in at least one model of the premises; i.e., possibly but not necessarily true), and 'impossible problems' (in which there is no model of the premises in which the conclusion is true; i.e., the conclusion is necessarily false).

In their first experiment, which used immediate inference tasks (in which conclusions are drawn from single quantified premises rather than from pairs of premises, as in syllogisms), the responses to possible problems under necessity instructions (that is, responses from subjects who were asked to determine which conclusions were necessarily true) showed a pattern which revealed two distinct types of possible problems. Logically, no conclusions to possible problems should be endorsed, since they are all possibly *but not necessarily* true (in model theoretic terms, supported by at least one but not all models of the premise). Seven of the twelve possible problems were endorsed by between 41% and 91% of subjects under necessity instructions, suggesting that many people had constructed an initial model for these problems which supported the conclusion but had failed to search for counterexamples (Evans *et al.* 1999: 1499). These problems were termed 'Possible Strong' problems. The other five possible problems were endorsed by between only 3% and 11% of subjects. These were termed 'Possible Weak' problems. There is no principled

© 2003. John Benjamins Publishing Company
All rights reserved

reason why people should have searched for counterexamples in Possible Weak problems but not in Possible Strong problems, so it is more likely that the initial model of the premise constructed for Possible Weak problems failed to support the putative conclusion.

Similar results were obtained in a syllogistic experiment (Evans *et al.* 1999: 1502). For example, 90% of subjects endorsed the following conclusion as necessary, even though it is only possible (that is, supported by at least one but not all models of the premises):

All A are B
Some B are not C
Therefore, Some A are not C

90% of subjects under possibility instructions also endorsed this conclusion. This suggested that the first model of the premises which most people constructed supported the conclusion, but that people who were instructed to determine whether this conclusion was necessary failed to look for a counterexample model. In contrast, the following syllogism, which is also possible (but not necessary) was endorsed by only 3% of subjects under necessity instructions. This suggested that the first model of the premises constructed by most people did not support the conclusion:

Some A are B
All C are B
Therefore, all C are A

Interestingly, rather than a similarly low proportion of subjects endorsing this conclusion under possibility instructions (as one would expect given that the most common initial model fails to support the conclusion), 43% endorsed it (correctly) as possible. This suggests that people are more likely to seek counterexamples in order to prove possibility than necessity.

Two different kinds of verificational (or counterexample) model are required in order to determine that a conclusion is necessary or possible. In the case of necessity, the initial model must support the conclusion and a reasoner should search for a counterexample to that initial model. That is, the object of the search is a model that does *not* support the conclusion. If such a model exists, the conclusion is not necessary; if there is no such model, the conclusion is verified. In the case of possibility, a conclusion can be endorsed if a single model is found that supports it, so if the initial model fails to support the putative conclusion a reasoner should search for a model that does support the

© 2003. John Benjamins Publishing Company
All rights reserved

conclusion. What Evans *et al.* (1999) demonstrated was that people are much more likely to search for an alternative model in order to endorse a conclusion as possible when the initial model *fails* to support the conclusion than they are to search for alternative models in order to endorse conclusions as necessary when the initial model *does* support the conclusion. They write,

> under instructions for Possibility, reasoners *do* seek alternative models of the premises if the first fails to support the premises [i.e., in Possible Weak problems], rather than simply responding No. This contrasts with the very high acceptance rates on Possible Strong problems under Necessity instructions, which suggest that rather little search for alternative models occurs when the first model identified confirms the conclusion (Evans *et al.* 1999: 1505).

The results reported by Evans *et al.* (1999) are in line with relevance theory. The search for counterexamples constitutes additional processing effort and will only occur if there is a strong likelihood of the additional effort being offset by additional contextual effects. It is to be expected, therefore, that additional processing effort may occur during the evaluation of conclusions as possible, since in everyday reasoning determining whether a state of affairs is possible typically generates significant contextual effects. Reasoners are rarely expected to determine instances of necessity, on the other hand, since rational communicators are usually explicit when it is important to know that a given states of affairs is necessary. People are not only used to having necessary states of affairs represented to them directly by their interlocutors, they are also less practised at determining when conclusions are necessary.

4. Belief bias effects

Belief bias is the term used in studies of syllogistic reasoning for the basing of responses on believability rather than on logical validity. In other areas of reasoning research it is known as 'memory cuing' or 'content effect'. In syllogistic reasoning, belief bias can introduce errors in two ways: (i) a logically valid conclusion may be rejected if it is unbelievable, and (ii) a logically invalid conclusion may be accepted if it is believable. The second case is more frequent, as the following table shows.

The response patterns for believable and unbelievable conclusions suggest that at least some people attempt to reason rather than simply accepting or rejecting a conclusion on the basis of (un)believability alone (see Newstead *et al.*

© 2003. John Benjamins Publishing Company
All rights reserved

Table 1. Overall percentage of conclusions (across three experiments) of Evans *et al.* (1983); compiled by Newstead *et al.* (1992: 262).

	Believable	**Unbelievable**
Valid	89	56
Invalid	71	10

1992: 266 for further experimental support for this claim). However, the results suggest that having reasoned to a certain extent, believability influences which conclusions are accepted by many people. These belief bias effects are uneven: a logically invalid conclusion is more likely to be accepted if it is believable (71%) than a valid conclusion is to be rejected if it is unbelievable (56%). Studies in which belief bias has been compared with belief-neutral control conditions (e.g., Evans and Pollard 1990; Newstead, Pollard, Evans, and Allen 1992) have shown that rather than accepting more believable conclusions, people accept fewer unbelievable conclusions compared with neutral cases. Thus belief bias is really belief *de*bias; "fallacies that would normally be made are withheld when the conclusion is unbelievable" (Evans *et al.* In press). This trend is reinforced by the fact that studies of belief bias prior to Evans *at al.* (In press) have investigated only Necessary (valid) syllogisms and Possible Strong syllogisms (in which the presented conclusion is possible but not necessary and in which the first model of the premises usually supports the putative conclusion). As Evans *et al.* note, since acceptance rates on Necessary and Possible Strong syllogisms are typically high, there is ample scope for improvement. In order to test for a belief bias effect due to believable conclusions, Possible Weak and Impossible syllogisms would have to be used, since acceptance rates on these are typically low.

The two most promising accounts of how belief bias operates (that is, the hypotheses compatible with most experimental evidence) are both based on mental models, but differ in terms of how people are assumed to reason when presented with unbelievable conclusions. Before looking at these theories and the experimental evidence, let us consider what predictions relevance theory makes in relation to evaluating syllogisms with believable or unbelievable conclusions.

4.1 Relevance theory

The relevance theory account assumes that people apply the same heuristic to reasoning as they do to utterance interpretation; that is, that people 'reason' in

© 2003. John Benjamins Publishing Company
All rights reserved

accordance with the principle of relevance. The putative conclusion of a syllogism (whether presented to the subject by the experimenter or produced by the subject) constitutes a source of new information for the subject of an experiment. According to relevance theory, a deduction based on the union of new information and old information may give rise to contextual effects (Sperber and Wilson 1986:108). Contextual effects come in three kinds: contextual implications, strengthening of existing assumptions, and contradiction with the consequent elimination of existing assumptions. If a conclusion is believable, it will achieve contextual effects by combining with existing assumptions and strengthening these. If a conclusion is unbelievable, it contradicts one or more of the subject's existing assumptions and can therefore achieve contextual effects only if it succeeds in eliminating an existing assumption.

According to relevance theory, a believable conclusion will give rise to a low level of contextual effects (a small modification to the subject's knowledge base) in the form of strengthening of some existing assumption, and so only warrants a small amount of processing effort. An attempt to falsify such a conclusion would result in additional processing effort with little likelihood of additional contextual effects. An unbelievable conclusion, on the other hand, would, if true, yield adequate contextual effects only if it succeeded in eliminating an existing assumption that it contradicts. Since the assumptions that are contradicted by unbelievable conclusions in belief bias experiments are ones that are strongly held by most subjects (for example, 'tigers are animals', 'cigarettes are addictive'), the level of contextual effects achieved by eliminating such an assumption would be very high (the modification of the subject's knowledge base would be great). Such a high level of contextual effects justifies additional processing effort in order to either falsify or validate an unbelievable conclusion. In particular, the validity of the new information must be verified in order for it to be held strongly (Nicolle 2000), which is required if new information is to eliminate any existing assumption which it contradicts. The relevance theory heuristic therefore predicts that logic, in the form of attempts to falsify initial conclusions, will play a larger role when conclusions are unbelievable than when they are believable. This prediction is born out by the experimental results presented in Table 1 above. In summary then, relevance theory predicts that people are likely to devote more processing effort to establishing the validity of unbelievable conclusions than to establishing the validity of believable conclusions, and the results from Possible Strong syllogisms bear this out.

Now, when reasoning through the construction and evaluation of mental models, a conclusion is recognised as valid when it can be shown that there is

© 2003. John Benjamins Publishing Company
All rights reserved

no model in which the conclusion and the premises are inconsistent. There are two ways in which to approach this task using mental models: positive and negative testing. A positive test requires every model of the premises to be consistent with the conclusion in order for the conclusion to be shown to be valid. A negative test searches for a model in which the premises are consistent with the logical *opposite* of the conclusion; failure to find such a model shows that the conclusion is valid. The choice between positive and negative testing is at the heart of the difference between the accounts of Newstead, Pollard, Evans and Allen (1992) and Klauer, Musch and Naumer (2000), both of which are discussed below.

4.2 Newstead, Pollard, Evans and Allen

Newstead *et al.* (1992) propose that subjects apply positive tests to putative conclusions, but only construct more than one model if the initial conclusion is unbelievable. If the initial conclusion is believable, then it is accepted and no further processing occurs (Newstead *et al.* 1992: 270). This explains the higher rate of logically correct responses in syllogisms with unbelievable conclusions. The following figure illustrates how relevance theory and Newstead *et al.*' s application of mental models theory would together determine people's reasoning processes with respect to syllogisms yielding believable and unbelievable initial conclusions.

Now, the model presented above predicts that there will be no difference dependent on the believability of conclusions in single model syllogisms. In a single model syllogism, the initial model will yield the correct conclusion; that is, there will be no alternative model that falsifies the conclusion drawn from the initial model. Therefore, whether a person searches for alternative models (with an unbelievable conclusion) or not (with a believable conclusion) will make no difference. This can be illustrated as follows. The syllogism below has a believable conclusion. If an initial model is constructed which represents only the identity relation, this will be compatible with the presented conclusion. Since the presented conclusion is believable, no further processing takes place, and the conclusion is accepted. Because this is a single model syllogism, the argument will be valid.

Valid single model syllogism with believable conclusion (Newstead *et al.* 1992: 264):

© 2003. John Benjamins Publishing Company
All rights reserved

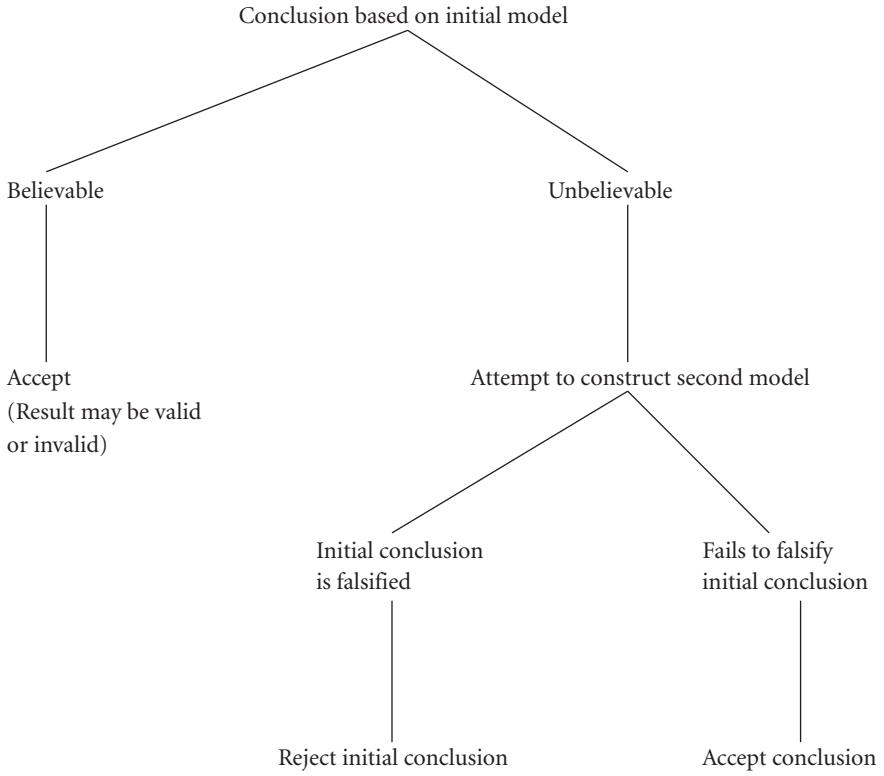

Figure 1. Reasoning processes in belief bias experiments based on Newstead et al. (1992)

All sparrows are haemopheds[7]
All haemopheds are birds
Therefore, all sparrows are birds

Initial model constructed representing only the identity relation:

[s] = [h] = [b]
[s] = [h] = [b]
[s] = [h] = [b]

If a syllogism of the same logical type is presented with an unbelievable conclusion, an alternative model of the premises (based on the subset relation) may be sought in an attempt to falsify the unbelievable conclusion. However, because this is a single model syllogism, there is no model that falsifies the presented conclusion.

© 2003. John Benjamins Publishing Company
All rights reserved

Valid single model syllogism with unbelievable conclusion:

All birds are haemopheds
All haemopheds are sparrows
Therefore, all birds are sparrows

Alternative model constructed to represent the subset relation:

[b] = [h] = s
[b] = [h] = s
[h] = s
s

Newstead *et al.* (1992: 271–275) found that there was in fact no interaction between logic and believability in single model syllogisms. With multiple model syllogisms, on the other hand, a subject who accepts a conclusion based on the first model constructed, simply because this is believable, risks accepting a conclusion that is not in fact valid. In the syllogism below, the presented conclusion is consistent with the identity model but is falsified by the subset model and the argument is therefore invalid. Newstead *et al.* (1992) predict that people will typically not attempt to falsify the presented conclusion, and will therefore accept an invalid argument.

Invalid multiple model syllogism with believable conclusion:

All birds are haemopheds
All haemopheds are sparrows
Therefore, all sparrows are birds

Initial model constructed representing only the identity relation:

[b] = [h] = [s]
[b] = [h] = [s]
[b] = [h] = [s]

Alternative model representing the subset relation (typically not constructed):

[b] = [h] = s
[b] = [h] = s
[h] = s
s

Conversely, Newstead *et al.* (1992) predict that people will attempt to falsify unbelievable conclusions. If the argument is valid (as in the example below) any attempt to find a falsifying model will fail, and a rational subject will therefore

© 2003. John Benjamins Publishing Company
All rights reserved

accept the conclusion on the basis of logic rather than believability. If the argument is invalid, a subject who attempts to falsify it may in fact do so, in which case the conclusion will be rejected. The uneven response pattern with respect to believable and unbelievable conclusions is therefore due entirely to multiple model syllogisms, as predicted by Newstead *et al.*' s (1992) application of mental models theory.

Valid multiple model syllogism with unbelievable conclusion (Newstead *et al.* 1992: 277):

> No animals are inhabitants of the island
> Some tigers are inhabitants of the island
> Therefore, some tigers are not animals

Finally, it should be noted that the results discussed in Newstead *et al.* (1992) present tendencies rather than absolute findings. For example, although fewer people made logical errors with unbelievable conclusions, the proportion of people who erroneously rejected valid arguments with unbelievable conclusions in Evans *et al.* (1983) was still over half. Not every subject in the experiments followed the relevance theory heuristic or manipulated the necessary mental models correctly. As Newstead *et al.* (1992: 283) conclude, performance in syllogistic reasoning tasks will probably only be fully explained by taking a number of theories into account.

4.3 Klauer, Musch and Naumer

Klauer, Musch and Naumer (2000) note that Newstead *et al.*' s failure to find a belief bias effect in single model syllogisms, in addition to being a null result, is at odds with the findings of Gilinsky and Judd (1994). Klauer *et al.* (2000) also found significant belief bias effects with single model syllogisms, as discussed below. They conducted a series of experiments using both single model and multiple model syllogisms in which subjects were given information about the base rate (the proportion of valid to invalid conclusions) in the experiment. Subjects were divided into high perceived base rate, medium perceived base rate, and low perceived base rate groups by alterations of the instructions. For example, subjects in the low perceived base rate groups were told that "on average every sixth of the twelve conclusions that you will see follows logically from the information given in that problem" (Klauer *et al.* 2000: 864). The medium perceived base rate was one half and the high perceived base rate was five sixths. It was found that although perceived base rate affected the responses

© 2003. John Benjamins Publishing Company
All rights reserved

it did not operate at the reasoning stage. That is, "[w]hen reasoners cannot deduce a confident evaluation, they use extra-logical clues such as base rate in coming to a decision" (Klauer *et al.* 2000: 878).

Klauer *et al.* (2000: 871–872) propose that if a conclusion is believable, people perform a positive test whereas if a conclusion is unbelievable people perform a negative test, and furthermore that naive reasoners rarely construct more than one model.[8] Thus, if presented with a believable conclusion, most subjects will attempt to construct a single model that integrates the premises and the conclusion, but if presented with an unbelievable conclusion, more subjects will attempt to construct a single model that integrates the premises and the logical negation of the conclusion. This is very similar to the Selective Processing Model of belief bias proposed by Evans (2000) (cited in Evans *et al.* In press), as illustrated in Figure 2. It also conforms to the relevance theory heuristic in that assessing unbelievable conclusions warrants additional processing effort because a negative test, even if only a single model is constructed, involves the additional processing effort of determining the logical negation of the conclusion.

Klauer *et al.* (2000: 871) also propose that successful model construction (constructing a model that is consistent with the premises and the conclusion or its negation) is taken as strong evidence for validity in single model syllogisms and as ambiguous evidence in multiple model syllogisms. In contrast, failure to construct a model that is consistent with the premises and the conclusion or its negation is always treated as providing ambiguous evidence.[9] Failure to construct a model that integrates the premises and the conclusion (or its negation) is taken as ambiguous evidence because subjects are assumed not to know whether to attribute their failure to the syllogism itself being invalid (or valid in the case of unbelievable conclusions), to semantic or other extra-logical difficulties inherent in the content of the premises, or to limitations in their own mental capacities.

The effect of belief bias noted in single model syllogisms by Gilinsky and Judd (1994) and Klauer *et al.* (2000) can be accounted for as follows in this theory. Some subjects assume that failure to construct a confirming model in a syllogism with a believable conclusion does not guarantee that the syllogism is invalid, since the failure may be due to extra-logical factors. Conversely, failure to construct a contradictory model in the case of a single model syllogism with an unbelievable conclusion does not guarantee the validity of the unbelievable conclusion. Thus, belief bias in single model syllogisms results when subjects fail to construct a confirming or falsifying initial model of the premises and the

© 2003. John Benjamins Publishing Company
All rights reserved

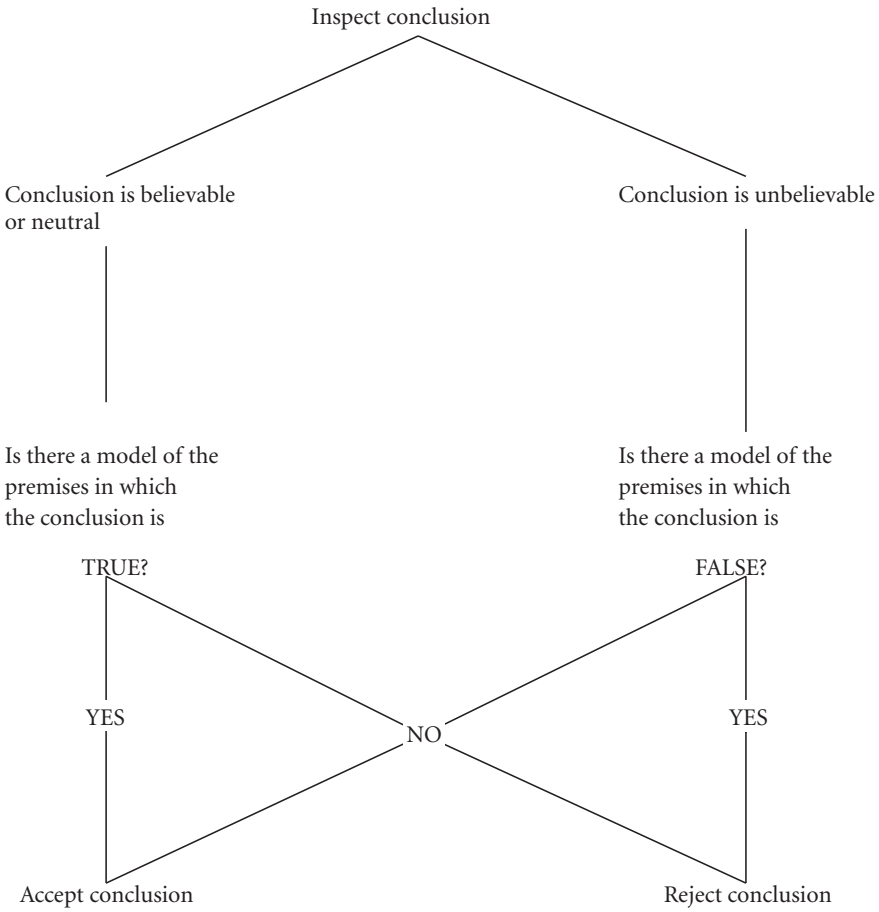

Figure 2. Reasoning processes in belief bias experiments based on the Selective Processing Model of Belief Bias (Evans 2000)

conclusion or its negation, but do not accept this failure as strong evidence. The reasoning stage of the decision process is only affected qualitatively by belief (believability prompts a positive test, unbelievability prompts a negative test), but the response stage in a situation where the logical evidence is perceived to be weak is influenced by extra-logical factors. These may include believability, perceived base rate, and the atmosphere effect (Begg and Denny 1969; Woodworth and Sells 1935), whereby a negative in the premises (*No*; *Some…not*) encourages a negative conclusion and an existential quantifier in the premises (*Some*; *Some…not*) encourages an existential quantifier in the conclusion,

© 2003. John Benjamins Publishing Company
All rights reserved

which in the single model syllogisms tested favours valid conclusions (Klauer *et al.* 2000: 875).

However, the results obtained by Klauer *et al.* (2000) in experiments using single model syllogisms do not demonstrate a large belief bias effect.[10] This may reflect a situation in which some subjects reason as Klauer *et al.* (2000) suggest: positive tests for believable conclusions, negative tests for unbelievable conclusions; but with a number of subjects reasoning in line with the model of Newstead *et al.* (1992): positive tests for all.

Even in multiple model syllogisms Klauer *et al.* (2000) assume that only one model is ever constructed, but it seems highly likely that in certain situations people can be encouraged to construct more than one model (as for example in the experiments reported here in Section 5, and in Bucciareli and Johnson-Laird 1999). If failure to construct a model is always treated (subjectively) as ambiguous evidence, relevance theory would predict that failure in situations involving an unbelievable conclusion is more likely to encourage further attempts to construct a falsifying model than failure with a believable conclusion is to encourage further attempts to construct a confirming model. This is because an unbelievable conclusion has the potential to yield a high level of contextual effects, but only if held with a strong degree of certainty — hence the justification for making the additional processing effort needed to establish logical validity. The results presented in the experiments discussed cannot be used to determine whether any subjects did construct additional models in such cases; this could potentially be determined by protocol experiments in which subjects are asked to think aloud whilst reasoning or to produce diagrams reflecting their reasoning strategies.

4.4 Discussion

I mentioned earlier that the use of Possible Strong syllogisms in these experiments (that is, syllogisms in which the first model constructed typically supports the putative conclusion, even though the argument is not in fact valid) allowed for a facilitating effect to be observed with unbelievable conclusions. In order to find evidence of belief bias due to acceptance of believable conclusions above the base rate of neutral controls, Impossible and Possible Weak syllogisms would have to be used. Evans *et al.* (In press) investigated Possible Weak syllogisms with believable, unbelievable and neutral conclusions, and found that "there is a significant increase in the normally low endorsement rates on these [i.e., Possible Weak] problems when the conclusion presented for

© 2003. John Benjamins Publishing Company
All rights reserved

evaluation is believable". They assume that people typically apply a positive test in such cases and attempt to find a model that supports the conclusion, but that this model is rarely the first one constructed. However, they suggest that when the conclusion is believable this either increases the relative availability of the supporting model (so the first model constructed is not one that falsifies the conclusion) or increases the relative amount of effort which people are willing to expend in order to find a supporting model. How the believability of a conclusion can affect the availability of a particular model (that is, change the order in which models of the premises are considered) is unclear, so the hypothesis that believability encourages the construction of additional models is more plausible. However, this is in direct contradiction to the relevance theory heuristic proposed above, in which believable conclusions warrant *less* processing effort than unbelievable ones.

What Evans *et al.* (In press) fail to note is that even if a search for an alternative model takes place, the initial model will already have falsified the presented conclusion. Their proposal amounts to a claim that a believable conclusion encourages people to respond not non-logically but counter-logically. What the relevance theoretic heuristic suggests, on the other hand, is that since believable conclusions warrant minimal processing effort, a proportion of subjects fail to reason at all; that is, some people simply accept believable conclusions without constructing even one model of the premises. Necessary and Possible Strong syllogisms have high acceptance rates anyway, so such non-reasoning responses merely follow the general trend and are unlikely to be noticed through statistical analysis. Possible Weak syllogisms, on the other hand, have low acceptance rates, so such non-reasoning responses will run counter to the general trend and therefore give rise to statistically significant differences between believable and unbelievable/neutral conclusions. Crucially, Evans *et al.* (In press) failed to include Impossible syllogisms in their study.

5. Modal auxiliaries and immediate inference tasks

In all the experiments reported so far, subjects were presented with a single conclusion that they were asked to evaluate. As we have seen, there is significant evidence that suggests that in such cases people perform either a positive or a negative test, depending on the nature of the putative conclusion, typically involving the construction of a single model. It is possible that different reasoning strategies may be employed when people are asked to produce their

© 2003. John Benjamins Publishing Company
All rights reserved

own conclusions to syllogisms (production tasks) or when people are presented with a choice of potential conclusions and asked to decide which, if any, are valid (multiple choice tasks). Bucciarelli and Johnson-Laird (1999) have shown that, under certain conditions, people can construct counter-example models as a method of reasoning, and this is what was proposed in Nicolle (1997) also. This study described multiple choice experiments in which the modal auxiliary *must* (and its opposite *can not*) was incorporated into the conclusions of immediate inference (single premise) tasks.

According to a relevance theoretic account of English modal auxiliaries (Groefsema 1995), an utterance of the form *must p* communicates that *p* is entailed by the set of assumptions that have a bearing on *p*. Informally, the set of assumptions that have a bearing on *p* consists of those assumptions which together determine the relevance of *p*. The contrast between *must* and the declarative form ('All A are B', etc.) on this account is not a matter of relative force or certainty; the crucial difference is that *must* introduces a conclusion based on inference rather than on observation or prior experience. Not only does the modal auxiliary *must* mean that a conclusion follows from the set of assumptions that have a bearing on the embedded proposition, but *must* also indicates that this proposition can be computed on the basis of deductive arguments taking these assumptions as premises. Hence the difference in interpretation between utterances (1) and (2) below:

(1) It must be hot in Djibouti at the moment.

(2) It will be hot in Djibouti at the moment.

(1) indicates that the proposition 'It is hot in Djibouti at the moment' has been determined on the basis of a deductive argument based on certain premises (for example, Djibouti has the highest average temperature of any permanently inhabited location, Djibouti is in the northern hemisphere, (1) is uttered during the month of August); (2) indicates that this proposition has been reached through observation or prior experience.

One interpretation of Groefsema's analysis of the modals is to treat these as exponents of what in relevance theory is termed 'procedural encoding' (Klinge 1993; Nicolle 1998a: 226–228, 1998b; Wilson and Sperber 1993). Procedural encoding functions by constraining the inferential computations performed over conceptual representations of a discourse. A modal auxiliary constrains the inferential computations performed over *p,* and can therefore be viewed as encoding procedural information. Wilson and Sperber (1993: 2) describe procedural encoding as providing information about the manipulation of

© 2003. John Benjamins Publishing Company
All rights reserved

conceptual representations, in contrast to conceptual encoding which provides information about the construction of conceptual representations. Mental models theory uses similar terminology when it states that reasoning and discourse comprehension both involve the construction and manipulation of mental models. Since propositional representations are interpreted with respect to mental models (Johnson-Laird 1983: 156), it is possible that procedural encoding, which influences the manipulation of propositional representations during discourse comprehension, might also influence the manipulation of mental models during reasoning.

In syllogisms and immediate inference tasks, the set of propositions that have a bearing on a conclusion consists of the premises plus the experimenters' instructions. The presence of *must* in conclusions to such tasks indicates that the conclusion is entailed by the premises and the experimenter's instructions. Now, this is no more than what is communicated by experimental instructions that emphasise logical validity, but the difference is that *must* encodes this information procedurally rather than conceptually. Procedural information is characterised as operating sub-consciously during the process of discourse comprehension, and it appears that reasoning through mental models is also largely a sub-conscious activity. On this view, if the presence of *must* in an utterance facilitates discourse comprehension, then the presence of *must* in a reasoning task should also facilitate the reasoning process.

5.1 Experimental results[11]

Two experiments were reported in Nicolle (1997), each of which was conducted under two conditions. In the first experiment subjects were presented with four premises in turn, each followed by four potential conclusions whose variables were ordered in the same way as those in the premise; in the second experiment the order of the variables in the conclusions was reversed. Each experiment was conducted under two different conditions. Under the non-modal condition, conclusions were presented as simple declaratives, 'All As are Bs', 'No As are Bs', etc. Under the modal condition conclusions contained modal auxiliaries, 'All As must be Bs', 'No As can be Bs', etc. In each experiment and condition subjects were instructed to decide which conclusions followed logically from each premise, given that the quantifier *Some* was defined as meaning 'At least one and possibly all'.

The interesting results (i.e., those in which subjects' performance differed in the two conditions) are summarised below.

© 2003. John Benjamins Publishing Company
All rights reserved

Table 2. Proportions of incorrect responses under standard and modal conditions in Nicolle (1997a)

Premise	False Conclusion	Condition 1 (%)	Condition 2 (%)
Some As are Bs	Some As are not Bs	100	38
Some As are not Bs	Some As are Bs	100	62
Some As are Bs	Some Bs are not As	90	53
Some As are not Bs	Some Bs are not As	87	53
Some As are not Bs	Some Bs are As	87	73

All of the subjects in the first experiment (in which the order of terms in the premises and conclusions was the same) treated *Some As are Bs* and *Some As are not Bs* interchangeably under the non-modal condition; that is, all subjects indicated that *Some As are Bs* entails *Some As are not Bs* and vice versa. However, this is a logical error since the quantifier *Some* was defined as 'at least one and possibly all'. Under the modal condition, only 38% of the subjects indicated that *Some As are Bs* entails *Some As must not be Bs* and 62% said that *Some As are not Bs* entails *Some As must be Bs*.

In the second experiment (in which the order of the terms in the conclusions was the reverse of that in the premises) almost all subjects under the non-modal condition reasoned that *Some Bs are not As* was entailed by both *Some As are Bs* and *Some As are not Bs*, whereas only half of the subjects under the modal condition made this logical error. Almost all of the non-modal condition subjects also reasoned (invalidly) that *Some As are not Bs* entailed *Some Bs are As*, compared with just under three quarters of the modal condition subjects.

These results demonstrate that modal wording in the conclusions of immediate inference tasks facilitates performance to varying degrees. The obvious questions are: (a) why does modal wording facilitate performance in these tasks? and (b) why does modal wording facilitate performance to different degrees with different tasks?

Nicolle (1997) analysed these results in terms of a modified version of mental models theory in which the following alterations to Johnson-Laird's (1983, 1986) account were proposed:

1. people construct a single, simple model of each premise with no potential arguments;
2. people draw a conclusion from this model but do not usually attempt to falsify it by constructing an alternative model of the premise;

© 2003. John Benjamins Publishing Company
All rights reserved

3. with modal wording in conclusions, however, more people construct an alternative (simple) model of each premise to test the initial conclusion.

The mental models which Nicolle (1997) proposed for the premises involving existential quantification are as follows (these models correspond to Euler circles and are numbered for future reference):

Some As are Bs:

1)	2)	3)	4)
a	a	a = b	a = b
a = b	a = b	a = b	a = b
a = b	a = b	a = b	a = b
b	a = b	b	a = b

Some As are not Bs:

1)	2)	5)
a	a	a
a = b	a = b	a
a = b	a = b	b
b	a = b	b

In the first experiment, both argument (i) (*Some As are Bs → Some As are not Bs*) and argument (ii) (*Some As are not Bs → Some As are Bs*) appear valid if only models 1 and 2 are considered. Since all subjects under the non-modal condition accepted these arguments, it is possible that these subjects only considered (one of) these models. Argument (i) is invalidated by model 3 or model 4, and argument (ii) is invalidated by model 5. The fact that 62% of modal condition subjects rejected (i) but only 38% of modal condition subjects rejected (ii) suggests that model 5 may be more difficult to construct than models 3 or 4.

In the second experiment, the order of the arguments in the premises and the conclusions was reversed. The three invalid arguments: (i) (*Some As are Bs → Some Bs are not As*), (ii) (*Some As are not Bs → Some Bs are not As*), and (iii) (*Some As are not Bs → Some Bs are As*) all appear valid if only model 1 is considered. (Argument (i) also appears valid given model 3, argument (ii) given model 5, and argument (iii) given model 2). Under the non-modal condition, most subjects (87% and 90%) accepted all three arguments, suggesting that these people only considered model 1. Arguments (i) and (ii) are both invalidated by model 2 and both were rejected by 47% of modal condition subjects[12] (argument (i) is also invalidated by model 4). Argument (iii) is invalidated by model 5, which the results of the first experiment suggested may be difficult to construct; this hypothesis is born out by the fact that only 27% of modal

© 2003. John Benjamins Publishing Company
All rights reserved

condition subjects managed to reject argument (iii).

From these results and analysis Nicolle (1997) concluded that the addition of the modal auxiliary *must* to the potential conclusions of an immediate inference task improves people's performance by prompting them to consider more than one simple mental model of the premise. The variation in the results is attributed to the various models being considered in a preferred order (corresponding, presumably, to the relative difficulty people have in constructing each model: the easiest model to construct being the first to be considered). It is therefore important to know if there is a usual order in which models are constructed (Evans and Over 1996: 159).[13]

The analysis of the results above suggests that virtually all of the subjects under the standard condition only constructed model 1 as a representation of either *Some* or *Some…not*. This is in line with the findings of Begg and Harris (1982: 617) who concluded that *Some* and *Some…not* are both interpreted as 'Some but not all' (see also Fisher 1981; and Newstead and Griggs 1983). Under the modal condition, performance was facilitated to varying degrees. Where the premise contained the quantifier *Some*, performance was facilitated by almost 50% when the conclusion was invalidated by model 2 and/or 4 and by 62% when the conclusion was invalidated by model 3 and/or 4, suggesting that model 3 might be slightly more frequently considered than model 2. Where the premise contained the negative existential quantifier *Some… not*, performance was facilitated by almost 50% when the conclusion was invalidated by model 2 and by only 38% and 27% in arguments invalidated by model 5, suggesting that model 2 is more often constructed than model 5. A tentative conclusion is that the models are typically considered in the order $1 > 3/2 > 4$ as representations of a premise containing *Some*, and $1 > 2 > 5$ as representations of a premise containing *Some… not*.

Previous studies of syllogistic reasoning, to which we shall turn shortly, have suggested that performance is linked to the order in which mental models of the premises are considered, and that performance can be improved by guiding people to create the 'best' (i.e., most informative) model *first* (see also Newstead and Griggs 1999). In everyday problem solving, the order in which models are constructed may be determined by practical utility (see Evans and Over 1996: 68). The presence of *must* in a conclusion, however, appears not to change the *order* in which models are considered; rather, it encourages the construction of *more than one* model, as subjects check that a modally worded conclusion is entailed by other members of the set of models which characterise the premise.

© 2003. John Benjamins Publishing Company
All rights reserved

The addition of modal wording only improved people's performance when the premise contained the quantifier *Some* or *Some…not*; performance was not facilitated when the premise contained either of the universal quantifiers *No* and *All*, where the error rate was at the level of chance for both conditions. Why should this be so? The only model corresponding to *No As are Bs* is model 5, so there is no scope for the construction of more than one model of a premise containing *No*. Model 5 seems to be rarely considered as a representation of *Some As are not Bs* or *Some Bs are not As*, so perhaps errors in the argument *No As are Bs → Some Bs are not As* are not surprising. However the premise *All As are Bs* can be represented either by the identity model (model 4) or the subset model (model 3). In the argument *All As are Bs → Some Bs are As*, errors result from a failure to realise that model 4 is a valid representation of the conclusion *Some Bs are As*. In the argument *All As are Bs → All Bs are As*, however, errors result from only considering the identity model. These are 'conversion errors' (see Newstead 1990 for a discussion).

Assuming that the occurrence of *must* in a conclusion encourages the construction of additional models representing the premises *Some* and *Some… not*, the question remains as to why *must* does not also encourage the construction of the subset model representing the premise *All* in addition to the identity model. The effect of different individual reasoning strategies may well be relevant here. Ford (1995) provides evidence from protocol studies that there are two different kinds of reasoners: those who reason 'verbally', relying on logical inference rules, and those who reason 'spatially' by manipulating models (although Ford rejects Johnson-Laird's account of mental models). Figures and analysis in Ford (1995: 45–52) suggest that significantly more verbal reasoners than spatial reasoners make conversion errors involving *All*. If Ford (1995) and Nicolle (1997) are correct, 'verbal' reasoners are unlikely to be affected by modal wording, since this appears to facilitate performance through encouraging the construction of additional mental models, which verbal reasoners do not use. 'Spatial' reasoners, on the other hand, may be affected by modal wording since they manipulate mental models when reasoning. If most of the conversion errors involving *All* are made by verbal reasoners, whose performance is unlikely to be affected by modal wording, this would explain why the presence of modal auxiliaries failed to facilitate performance in immediate inference tasks where the premise contained the quantifier *All*. The prior question of why verbal reasoners should make more conversion errors from premises with *All* than spatial reasoners do remains to be answered.

© 2003. John Benjamins Publishing Company
All rights reserved

The central claim in Nicolle (1997) was that people typically do not attempt to falsify conclusions drawn on the basis of an initial model of the premises unless encouraged to do so by the inclusion of modal auxiliaries in potential conclusions. The claim that people do not usually attempt to construct an alternative model of the premise in order to test an initial conclusion is in line with the findings of Newstead, Handley and Buck (1999), Evans, Handley, Harper and Johnson-Laird (1999) and Evans, Handley and Harper (In press), who conclude that, whilst people are capable of searching for additional models, they do not necessarily do so spontaneously.

Newstead *et al.* (1999) assume that subjects' reflective responses provided an accurate picture of their reasoning processes. In one experiment, subjects were asked to solve a syllogism and immediately indicate, by ticking a list of potential conclusions, how many answers they had considered before coming to a decision. It is debatable, in particular considering that subjects were required to solve 24 syllogisms in turn, whether these people actually knew or remembered which conclusions they had considered for each task. As Newstead *et al.* admit, subjects might forget some of the intermediate stages involved in their reasoning. In another experiment reported in the same paper, one group of subjects were required to draw diagrams representing the syllogism, and to base their conclusions on these, whilst a control group was asked simply to solve the syllogisms. Newstead *et al.* found that there was no correlation between the number of diagrams constructed and success in solving multiple model syllogisms (that is, syllogisms in which more than one model must be constructed to ensure a correct response). However, this conclusion assumes (a) that the diagrams produced were accurate reflections of subjects' mental representations, and (b) that the conscious process of manipulating diagrams accurately reflects the sub-conscious manipulation of mental models. Furthermore, if Ford (1995) is correct, even if the diagrams drawn by spatial reasoners (who manipulate models) provide an accurate reflection of their mental processes, the same may well not be true for verbal reasoners. In contrast, the methodology of Nicolle (1997) provides (indirect) access to sub-conscious reasoning processes, rather than assuming that these are the same as the conscious processes of recall and diagram construction.

© 2003. John Benjamins Publishing Company
All rights reserved

6. Conclusion

All of the experiments discussed here provide support for the mental models theory account of reasoning with quantified statements. However, they also demonstrate that although people may use mental models as part of the reasoning process, the strategy that is typically employed in drawing conclusions from models is not that which is guaranteed to lead to valid conclusions. Rather than construct counter-example models in order to test an initial (or putative) conclusion, most people reason on the basis of only a single model (if indeed thay 'reason' at all). This is the reasoning strategy that I have suggested is employed during discourse comprehension according to relevance theory; that is, people apply the same deductive processes in the same way in logical reasoning and in utterance interpretation. This in turn suggests that people generally act 'rationally' in performing logical reasoning tasks even when they do not act logically.

In Sections 3 and 4 we saw that the reasoning process can be affected by the type of task being performed (determination of possibility as opposed to necessity) and by the believability of the putative conclusions. Performance may be facilitated when people perform 'negative tests' in which they attempt to validate the negation of the presented conclusion. Performance may also be facilitated by the addition of modal wording to conclusions in multiple choice tasks where at least one of the premises contains an existential quantifier. In these cases, improved success rates result from subjects having considered more than one model.

Both mental models theory and relevance theory receive support from this analysis, but the fact that mental models theory provides such a promising account of the mechanisms underlying reasoning with syllogisms, and that it is compatible with the predictions of relevance theory, means that relevance theory as it stands is in need of revision. Sperber and Wilson assume "that there is a set of deductive rules which are spontaneously brought to bear in the deductive processing of information" (Sperber and Wilson 1986: 85). The deductive rules posited include *and-elimination*, *modus ponendo ponens* and *modus tollendo ponens*. What the discussion in this paper suggests, however, is that people typically do not reason through employing a set of deductive rules, for if that were the case, we should not expect to see the systematic patterns of error in reasoning that we have seen. Human reasoning often arrives at logically valid conclusions, but the heuristic employed to reach these conclusions typically does not involve deductive rules. Relevance theory is perfectly compatible

© 2003. John Benjamins Publishing Company
All rights reserved

with the mental models theory of human reasoning, and can even help explain the way in which models are employed. If mental models theory and relevance theory are integrated in the way in which I propose, this may also contribute to the development of the distinction between conceptual and procedural encoding in relevance theory, as suggested in Section 5. Procedural encoding describes the linguistic encoding of information that constrains the computations performed over (or manipulation of) conceptual representations. Experimental research into the construction and manipulation of mental models has the potential to shed light on the construction and manipulation of conceptual representations of situations and entities, and hence to clarify the distinction between conceptual and procedural information.

Notes

* I am grateful to the participants at the Linguistics Association of Great Britain Workshop on Experimental Pragmatics held at the University of Luton, 10 September 1998, to Jonathan Evans for constructive comments on an earlier draft of this paper, and in particular to Stephen Newstead for introducing me to syllogistic reasoning and providing me with much valuable advice.

1. This is consistent with the suggestion in Begg and Harris (1982) that people treat syllogisms as obscurely stated attempts to communicate.

2. Relevance theory here is to be distinguished from Evans' (1989) use of the same term to refer to his theory concerning what determines the focus of subjects' attention during reasoning.

3. Recent work using a mental models approach (Evans 2000; Evans, Handley, and Harper, In press; Klauer, Musch, and Naumer 2000) has suggested an alternative heuristic when certain conclusions are presented for evaluation. According to this heuristic, a model of the premises is constructed and compared with the negation of the presented conclusion; if the initial model of the premises is consistent with the negation of the presented conclusion, then the conclusion has been falsified. This is discussed in greater detail below.

4. Preliminary experiments into the interaction between modal auxiliaries and believability of conclusions suggests that belief bias cancels out the effect of modal auxiliaries; that is, the (un)believability of a conclusion has a greater effect on people's responses than logical necessity overtly signaled through modal auxiliaries.

5. In the notation of Johnson-Laird and Byrne (1991) these two 'simple' models would be represented by a single compound model in which the dots indicate the possible existence of entities not represented:

[a] = b
[a] = b
. . .

© 2003. John Benjamins Publishing Company
All rights reserved

6. The terms in the syllogisms and immediate reasoning tasks used in some of the following experiments took the form of randomly selected letters of the alphabet. The results of these experiments should be replicable with alternative arguments, given that no significant difference was found between using letters and realistic categories (occupations and hobbies) in syllogistic reasoning tasks (Newstead 1989: 83).

7. An anonymous referee has pointed out that because 'haemopheds' is an invented word, subjects in such an experiment will not know how this term is supposed to be defined, and will therefore have no opinion about whether the premises really support the conclusion. To overcome this problem, in experiments of this kind subjects are told to treat the premises as true. If a known term had been used in the premises (for example 'seed-eaters'), then subjects' responses would have reflected their judgements about the validity of the premises, which was not the object of study for these experiments. It is also worth mentioning that the subjects in all of the experiments reported here had no training in formal logic, so the notion of logical validity (as opposed to contingrent truth) was explained in the instructions accompanying each task.

8. They write: "participants perform positive tests for syllogisms with believable conclusions, whereas they frequently perform negative tests in the case of unbelievable conclusions" (Klauer et al. 2000: 871–872). This use of 'frequently' in relation to negative tests is not quantified.

9. This is objectively true, but for naive reasoners to use this heuristic they would have to know when a syllogism is a single or multiple model one (a possibility mentioned by Evans *et al.* 1999: 1507). Since naive reasoners do not already have access to this information, they would have to determine it by attempting to construct additional models for each syllogism, which is exactly what Klauer *et al.* assume does NOT happen. However, if we accept that ALL successful model construction is accepted as strong evidence, then we can account for the larger effect of belief bias with believable than with unbelievable conclusions in multiple model syllogisms as follows. A single successful positive test (applied in situations with believable conclusions) does not guarantee the validity of the conclusion, but a single successful negative test (applied in situations with unbelievable conclusions) guarantees that the conclusion is invalid. Conversely, failure in a positive test guarantees that the conclusion is invalid, but failure in a negative test does not guarantee that the conclusion is valid. Thus success is more important in tests involving unbelievable conclusions and failure is more important with believable conclusions, but since failures are treated as ambiguous evidence, more decisions involving believable conclusions will be decided using criteria other than logic.

10. Klauer *at al.*' s study 5 in fact yielded results that were consistent with the null finding observed in Newstead *et al.* (1992) and so was repeated as an internet study (study 6) to obtain a larger sample.

11. Since the experiments described here have been reported in detail in Nicolle (1997), I will merely summarise the methodology employed and results whilst discussing the theoretical implications in greater detail.

© 2003. John Benjamins Publishing Company
All rights reserved

12. Although the percentages are identical, not all the people who rejected argument (i) also rejected argument (ii), even though there was a considerable overlap between these two groups. Detailed within-subject data are unavailable at this time.

13. Against the possibility that people construct and manipulate more than one model simultaneously, Mynatt, Doherty, and Dragan (1993) have argued that people can only entertain one mental model of a state of affairs at a time (cf. Evans and Over 1996: 66–68).

References

Begg, I. 1987. "Some". *Canadian Journal of Psychology* 41: 62–73.

Begg, I. and Denny, P. 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 Behavior* 21: 595–620.

Bucciarelli, M. and Johnson-Laird, P. N. 1999. "Strategies in syllogistic reasoning". *Cognitive Science* 23: 247–303.

Ceraso, J. and Provitera, A. 1971. "Source of error in syllogistic reasoning". *Cognitive Psychology* 2: 400–410.

Dascal, M. 1987. "Language and reasoning: Sorting out sociopragmatic and psychopragmatic factors". In B. W. Hamill, R. C. Jernigan, and J. C. Boudreaux (eds), *The Role of Language in Problem Solving, Vol. II*. Amsterdam: North Holland, 183–197.

Dickstein, L. S. 1978. "The effect of figure on syllogistic reasoning". *Memory and Cognition* 6: 76–83.

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. 1989. *Bias in Human Reasoning: Causes and Consequences*. Hove, UK: Lawrence Erlbaum.

Evans, J. St. B. T. 2000. "Thinking and believing". In J. Garcia-Madruga and C. González-Labra (eds), *Mental Models in Reasoning*. Madrid: UNED: 41–55.

Evans, J. St. B. T., Handley, S. J., Harper, C. N. J., and Johnson-Laird, P. N. 1999. "Reasoning about necessity and probability: A test of the mental model theory of deduction". *Journal of Experimental Psychology: Learning, Memory and Cognition* 25: 1495–1513.

Evans, J. St. B. T., Handley S. J., and Harper C. N. In press. "Necessity, possibility and belief: A study of syllogistic reasoning". *Quarterly Journal of Experimental Psychology*.

Evans, J. St. B. T., Newstead, S. E., and Byrne, R. M. J. 1993. *Human Reasoning: The Psychology of Deduction*. Hove, UK: Lawrence Erlbaum.

Evans, J. St. B. T. and Over, D. E. 1996. *Rationality and Reasoning*. Hove: Psychology Press.

Evans, J. St. B. T. and Pollard, P. 1990. "Belief bias and problem complexity in deductive reasoning". In J. P. Caverni, J. M. Fabre and M. Gonzales (eds), *Cognitive Biases*. Amsterdam: North-Holland, 131–154.

© 2003. John Benjamins Publishing Company
All rights reserved

Fisher, D.L. 1981. "A three-factor model of syllogistic reasoning: The study of isolable stages". *Memory and Cognition* 9: 496–514.

Ford, M. 1995. "Two modes of mental representation and problem solution in syllogistic reasoning". *Cognition* 54: 1–71.

Girrotto, V., Mazzocco, A., and Tasso, A. 1997. "The effect of premise order in conditional reasoning: A test of the mental model theory". *Cognition* 63: 1–28.

Groefsema, M. 1995. "*Can*, *may*, *must* and *should:* A relevance theoretic account". *Journal of Linguistics* 31: 53–79.

Hardman, D.K. and Payne, S.J. 1995. "Problem difficulty and response format in syllogistic reasoning". *Quarterly Journal of Experimental Psychology* 48A: 945–975.

Johnson-Laird, P.N. 1983. *Mental Models.* Cambridge: Cambridge University Press.

Johnson-Laird, P.N. 1986. "Reasoning without logic". In T. Myers, K. Brown, K., and B. McGonigle (eds), *Reasoning and Discourse Processes.* London: Academic Press, 13–49.

Johnson-Laird, P.N. 1995. "Inference and mental models". In S.E. Newstead and J. St. B.T. Evans (eds), *Perspectives on Thinking and Reasoning.* Hove: Lawrence Erlbaum, 115–146.

Johnson-Laird, P.N. and Bara, B. 1984. "Syllogistic inference". *Cognition* 16: 1–61.

Johnson-Laird, P.N. and Byrne, R.M.J. 1991. *Deduction.* Hove: Lawrence Erlbaum.

Johnson-Laird, P.N., Byrne, R.M.J., and Schaeken, W. 1992. "Propositional reasoning by models". *Psychological Review* 101: 734–739.

Klauer, K.C. and Oberauer, K. 1995. "Testing the mental model theory of propositional reasoning". *Quarterly Journal of Experimental Psychology* 48A: 671–687.

Klauer, K.C., Musch, J., and Naumer, B. 2000. "On belief bias in syllogistic reasoning". *Psychological Review* 107: 852–884.

Klinge, A. 1993. "The English modal auxiliaries: From lexical semantics to utterance interpretation". *Journal of Linguistics* 29: 315–357.

Lukasiewicz, J. 1957. *Aristotle's Syllogistic from the Perspective of Modern Formal Logic*, 2nd edition. Oxford: Oxford University Press.

Mynatt, C.R., Doherty, M.E., and Dragan, W. 1993. "Information relevance, working memory and the consideration of alternatives". *Quarterly Journal of Experimental Psychology* 46A: 759–778.

Newstead, S.E. 1990. "Conversion in syllogistic reasoning". In K. Gilhooly, M.T.G. Keene, R. Logie, and G. Erdos (eds), *Lines of Thought: Reflections on the Psychology of Thinking, Vol 1.* London: John Wiley 73–84.

Newstead, S.E. 1995. "Gricean implicatures and syllogistic reasoning". *Journal of Memory and Language* 34: 644–664.

Newstead, S.E. and Griggs, R.A. 1999. "Premise misinterpretation and syllogistic reasoning". *Quarterly Journal of Experimental Psychology* 52A: 1057–1075.

Newstead, S.E., Handley, S.J., and Buck, E. 1999. "Falsifying mental models: Testing the predictions of theories of syllogistic reasoning". *Memory and Cognition* 27: 344–354.

Nicolle, S. 1997. "Relevance and the effect of modal auxiliaries in logical reasoning tasks". In P. Weingartner, G. Schurz and G. Dorn (eds), *The Role of Pragmatics in Contemporary Philosophy: Contributions of the Austrian Ludwig Wittgenstein Society, Vol. 5.* Kirchberg am Wechsel: Österreichische Ludwig Wittgenstein Gesellschaft, 694–700.

Nicolle, S. 1998a. "*Be going to* and *will*: A monosemous account". *English Language and Linguistics* 2: 223–243.

© 2003. John Benjamins Publishing Company
All rights reserved

Nicolle, S. 1998b. "A relevance-theoretic perspective on grammaticalization". *Cognitive Linguistics* 9: 1–35.
Nicolle, S. 2000. "Communicated and non-communicated acts in relevance theory". *Pragmatics* 10: 233–245.
Polk, T. A. and Newell, A. 1995. "Deduction as verbal resoning". *Psychological Review* 102: 533–566.
Rips, L. J. 1994. *The Psychology of Proof: Deductive Reasoning in Human Thinking.* Cambridge, MA and London: The MIT Press.
Sperber, D. and Wilson, D. 1986. *Relevance: Communication and Cognition.* Oxford: Basil Blackwell.
Wetherick, N. E. and Gilhooly, K. 1990. "Syllogistic reasoning: Effects of premise order". In K. Gilhooly, M. T. G. Keene, R. Logie and G. Erdos (eds), *Lines of Thinking: Reflections on the Psychology of Thought, Vol 1.* London: John Wiley, 99–108.
Wilson, D. and Sperber, D. 1993. "Linguistic form and relevance". *Lingua* 90: 1–25.
Woodworth, R. S. and Sells, S. B. 1935. "An atmosphere effect in syllogistic reasoning". *Journal of Experimental Psychology* 18: 451–460.

Author's address

Dr. Steve Nicolle
P. O. Box 12
Kwale, KENYA

Email: steve-alison_nicolle@sil.org

About the author

Steve Nicolle (PhD in Linguistics) is a Linguistics Consultant with SIL International, based in Kenya. Before working in Kenya he taught at Middlesex University in London, and has published on tense and aspect, phatic communication, grammaticalisation, experimental pragmatics, Bantu languages and ethnobotanical taxonomies.

© 2003. John Benjamins Publishing Company
All rights reserved