

Reasoning About Necessity and Possibility: A Test of the Mental Model Theory of Deduction

Jonathan St. B. T. Evans, Simon J. Handley, and
Catherine N. J. Harper
University of Plymouth

Philip N. Johnson-Laird
Princeton University

This article examined syllogistic reasoning that differs from previous research in 2 significant ways: (a) Participants were asked to decide whether conclusions were possible as well as necessary, and (b) every possible combination of syllogistic premises and conclusions was presented for evaluation with both single-premise (Experiment 1) and double-premise (Experiment 2) problems. Participants more frequently endorsed conclusions as possible than as necessary, and differences in response to the 2 forms of instruction conformed to several predictions derived from the mental model theory of deduction. Findings of Experiments 2 and 3 showed that some fallacies are consistently endorsed and others consistently resisted when people are asked to judge whether conclusions that are only possible follow necessarily. This finding was accounted for by the computational implementation of the model theory: Fallacies are made when the *first* mental model of the premises considered supports the conclusion presented.

Traditional applications of logic concern the validity of arguments, that is, proving that some conclusion is *necessary* given some premises. However, in everyday reasoning it may be just as important to decide whether some proposition is *possible* in light of the given information. Inferences of possibility occur whenever rules and regulations constrain a person's behavior rather than determine what it must be. For example, students choosing a degree program within a modular course structure will have many degrees of freedom but will have to respect constraints owing to timetabling restrictions, availability of teaching staff, prerequisite and corequisite relations between modules, and so on. Their decision making here involves inferring what is possible and then deciding between the possibilities identified. Most of the rules by which people live their lives in society, including criminal law, operate in a similar way. It is rarely determined by such rules that people must follow a particular course of action, but they are frequently constrained to act within a set of legal possibilities.

The psychological study of deductive reasoning has been steadily increasing in recent years, and there are now many hundreds of experimental studies reported in the literature (see Evans, Newstead, & Byrne, 1993, for a review). In view of the above remarks, however, there is a curious limitation

in almost all studies reported to date: They ask participants only to decide if a conclusion is necessary given some premises. There have been almost no studies prior to that reported in the present article that ask participants to decide what is or is not possible given some assumptions.

In logic, a deductive inference is valid if its conclusion must be true given that its premises are true. Conclusions in standard logic state what is the case, but in "modal" logic they can in addition state what is possibly the case, or what is necessarily the case. Most psychological experimental investigations of reasoning have considered only conclusions about what is the case. Typically, the participants are asked to evaluate the validity of a given conclusion (the evaluation task) or else to generate their own valid conclusion (the production task). Most of the studies have been based on problems drawn from the propositional calculus, which concerns the logic of negation and such connectives as *if*, *or*, and *and*, or on classical syllogisms of the sort devised by Aristotle. There have been few studies of modal reasoning (but see Bell & Johnson-Laird, 1998; Galotti, Baron, & Sabini, 1986; Osherson, 1976).

In the research reported here, we examine inferences about what is possible as well as what is necessary within the syllogistic reasoning paradigm. The research is motivated by the mental model theory of deductive reasoning from which we are able to derive some general predictions about necessity and possibility in deductive reasoning. We start with a brief account of that theory.

Mental Model Theory of Deduction

The reason that so much research on standard deduction has been carried out is because many cognitive scientists believe that the ability to reason deductively is central to human intelligence. In fact, people make many logical errors on these tasks and exhibit systematic biases (Evans, 1989), a

Jonathan St. B. T. Evans, Simon J. Handley, and Catherine N. J. Harper, Department of Psychology, University of Plymouth, Plymouth, United Kingdom; Philip N. Johnson-Laird, Department of Psychology, Princeton University.

This study was supported by a research grant from the Economic and Social Research Council of the United Kingdom (R000221742).

Correspondence concerning this article should be addressed to Jonathan St. B. T. Evans, Center for Thinking and Language, Department of Psychology, University of Plymouth, Plymouth PL4 8AA United Kingdom. Electronic mail may be sent to J.Evans@plym.ac.uk.

finding which has led to much debate about human rationality (see Evans & Over, 1996; Manktelow & Over, 1993). However, most psychologists in the area would agree that untrained reasoners do show some basic deductive competence: For example, they can distinguish valid inferences from fallacies at well above chance rates. Accounting for such competence has, however, led to a protracted, sometimes heated, and as-yet unresolved debate between two theoretical camps. The debate is focused on whether deductive reasoning is achieved by the use of mental rules or mental models (see Evans et al., 1993, chapter 3, for detailed coverage of the debate).

Rule-based theories derive from so-called natural deduction, a method of formalizing logic that is intuitively plausible and that postulates rules of inference for each logical term. The most durable of these have proved to be the systems devised by Rips (1983, 1994) known as PSYCOP, and the system developed by Braine and O'Brien (Braine, 1978; Braine & O'Brien, 1991; Braine, Reiser, & Rumin, 1984). However, we provide little discussion of these theories in the present article for two reasons. First, rule theorists have largely confined their theoretical and experimental work to the study of propositional reasoning and have had rather little to say about syllogistic inference (but see some discussion of reasoning with quantifiers by Rips, 1994). Second, the rule theories as published provide only rules for valid inferences and hence can account for necessary but not possible inferences. We defer consideration of how rule theories might be extended to the explanation of possible inference until the General Discussion.

The account of human deduction in terms of mental model theory was devised by Johnson-Laird (1983; Johnson-Laird & Byrne, 1991). In this approach, people do not prove conclusions syntactically by applying valid inference rules. Rather, their deductions are based on grasping a semantic principle, namely, that a conclusion is valid if there is no model of the premises that exclude it. The model theory includes psychological constraints that do not apply to formal semantic methods such as truth-table analysis. In particular, it is supposed that people focus on situations in which the premises would be true and represent such situations selectively.

The mental model theory was originally devised as an account of syllogistic reasoning (Johnson-Laird & Bara, 1984) and in its general form supposes that deduction requires three stages:

1. *Model formation.* The reasoner forms an initial model from the premises.
2. *Conclusion formation.* The reasoner derives a putative conclusion from the model that is informative (e.g., not a repetition of a premise).
3. *Conclusion validation.* The reasoner searches for a counterexample, that is, a model in which the premises are true but the conclusion is false. If no such model is found, the conclusion is valid.

A large experimental program of research has been carried out by mental model theorists who claim support for the predictions of their theory (see Johnson-Laird & Byrne, 1991). For example, Johnson-Laird and Bara (1984) ran a

large-scale study of syllogistic reasoning using production task methodology. All possible premise pairs of the classical syllogistic form (see below) were presented, and participants were asked to derive a conclusion. The authors analyzed these premise pairs according to whether they were consistent with one, two, or three models. As predicted, error rates were higher on multimodel problems. The psychological principle here is that consideration of multiple models places a strain on working-memory capacity. In particular, people often commit fallacies because they fail to find a counterexample model at Stage 3, even though such a model exists.

Many specific findings claimed in support of the model theory have, however, been disputed by rule theorists (see Evans et al., 1993). Evans and Over (1996, 1997) argued that it is very difficult to decide between the two approaches in most of these arguments as neither theory is fully specified. For this reason, we present here an attempt to test some predictions that follow from only the most general principles of the model theory and that do not rely on any supplementary assumptions to account for the specific tasks. Nor do they rest on any assumptions about how particular connectives or quantifiers are represented. These predictions concern what people will do when asked to make judgments of necessity and possibility and are derived below.

Possible and Necessary Inference

In modal logic, a distinction is drawn among truths, necessary truths, and possible truths. A statement is true if it is a matter of fact; a statement is necessary if it *must* be true; and a statement is possible if it *may* be true. Consider the following three arguments based on universal premises:

All artists are beekeepers,
Lisa is an artist,
Therefore, it necessary that Lisa is a beekeeper. (1)

All artists are beekeepers,
Lisa is a beekeeper,
Therefore, it is possible Lisa is an artist. (2)

All artists are beekeepers,
Lisa is an artist,
Therefore, it is possible that Lisa is not a beekeeper. (3)

Argument 1 is a *modus ponens* inference in which the modal conclusion states a necessity. Given that the premises are true, then indeed it is necessary that Lisa is a beekeeper, and so the conclusion is valid. Argument 2 is also valid, but any stronger conclusion, such as "It is necessary that Lisa is an artist," would be invalid, because the premises are consistent with a situation in which there are beekeepers who are not artists, and so Lisa may be a beekeeper who is not an artist. Argument 3 is invalid, because as Argument 1 shows, Lisa is necessarily a beekeeper given the premises.

In this article, we refer to these three sorts of argument as Necessary, Possible, and Impossible. Standard deductive reasoning instructions ask people to make judgments about the validity of factual conclusions. If people are asked to decide whether it is necessary that X, where X is an assertion, then they should endorse Necessary conclusions and reject both those that are Possible and Impossible. If

people are asked instead to decide whether it is possible that X, then they should endorse Necessary conclusions and Possible conclusions and reject only Impossible conclusions. The ability to discriminate Possible from Impossible conclusions is neglected in normal psychological research on reasoning.

Consider the three general stages of the model theory defined earlier. In the first two stages, the reasoner constructs a single model of the premises that supports a provisional conclusion. This conclusion can be reported as possible without any further reasoning. The conclusion validation stage (search for counterexamples) only comes into play when people are asked to prove that the putative conclusion is necessary. Hence, according to the model theory, possible inference should be easier than necessary inference. Our first prediction (P) therefore is as follows:

P1. People will be more willing to endorse conclusions as possible than as necessary.

The model theory provides some further clear predictions about the relative difficulty of different types of problems in this paradigm. Consider the three basic types of problems from a model theory perspective:

Necessary problems: the conclusion is true in all models of the premises.

Possible problems: the conclusion is true in at least one model of the premises.

Impossible problems: the conclusion is true in no model of the premises.

From this analysis we derive and test in this article two specific predictions:

P2. It will be easier to decide that a conclusion is possible if it is also necessary. Specifically, we predict more endorsements of possibility for Necessary than for Possible problems.

P3. It will be easier to decide that a conclusion is not necessary if it is also not possible. Specifically, we predict that more Possible than Impossible problems will be endorsed as necessary.

P2 follows because it is only necessary to discover one model supporting the conclusion in order to infer that it is possible. On Necessary problems, all models support the conclusion, whereas on Possible problems there is at least one model that does not. Hence, reasoners are at risk of thinking first of a model that does not support the conclusion in the latter case. P3 follows because a conclusion is not necessary when there is at least one model that does not support it. On Impossible problems, there is no model of the premises that supports the conclusion, whereas on Possible problems there is at least one that does. Hence on the latter problems reasoners are at risk of first thinking of a model that supports the conclusion.

We report two experiments designed to test our a priori hypotheses and a third that addresses issues arising from the findings of these studies. In Experiment 1, we tested the above predictions using a simpler task than full syllogistic reasoning. The task involves giving people a single premise of the kind used in syllogisms from which they must draw an inference; this is usually referred to as the *immediate*

inference task (see Newstead & Griggs, 1983). Experiment 2 involves full syllogistic reasoning and to our knowledge is the first to ask people to evaluate every possible conclusion with regard to every possible pair of premises even for validity (Necessity) judgments. Both experiments included groups required to make judgments of Possibility as well as of Necessity of conclusions.

Experiments 1 and 2 were designed together and run on the same participants in the same sessions. Owing to the large difference in scale between the two experiments, however, counterbalancing of the order of the experiments seemed inappropriate. It was also undesirable to ask the same participant to make judgments of both necessity and possibility. The procedure adopted was as follows. Participants were allocated into Necessity and Possibility groups and run through the short task of Experiment 1 using the appropriate form of instruction. Each participant then performed the large task required by Experiment 2 using the same form of instruction allocated to them in Experiment 1. The groups were also subdivided in Experiment 2 according to the order of terms in the conclusions they evaluated (see *Method* section of Experiment 2 for details). We also report a third experiment that replicates some of the key findings of Experiment 2 using a much smaller subset of syllogisms and simpler method of problem presentation. Experiment 3 also provides a check that the results of Experiment 2 were not influenced by the prior experience of taking part in Experiment 1.

Experiment 1: Immediate Inference

Syllogisms are formed by combining two premises and a conclusion, each of which can have one of only four quantified forms known traditionally by the abbreviations A, E, I, and O and described as "moods." These are as follows:

A	Universal affirmative	All X are Y
E	Universal negative	No X are Y
I	Particular affirmative	Some X are Y
O	Particular negative	Some X are not Y

In Experiment 1, we looked only at the immediate inferences that people may draw when one of these statements acts as a premise and the other as a conclusion. Because the order of terms can be exchanged, there are seven possible conclusions (excluding the repetition of the premise) to be considered for each of the four possible premises.

As with syllogisms, immediate inference problems can be classified into three types, where there is no doubt about the existence of entities referred to by the various terms:

Necessary. The conclusion statement must be true given that the premise is true. Of the 28 problems, 6 are of this kind.

Possible. The conclusion statement might be true given that the premise is true. Of the 28 problems, 12 are of this kind.

Impossible. The conclusion statement cannot be true given that the premise is true. Of the 28 problems, 10 are of this kind.

Note that logically any conclusion that is necessary is also possible. The problems we classify as Possible are in fact ones in which the conclusion is possible but not necessary.

To demonstrate the method, we consider an example of what can be inferred from the premise "All A are B." Each of the following possible conclusions can be presented in turn for evaluation:

1. No A are B -
2. Some A are B
3. Some A are not B
4. All B are A
5. No B are A
6. Some B are A
7. Some B are not A

Conclusions 2 and 6 are necessary; 4 and 7 are possible (but not necessary), whereas 1, 3, and 5 are impossible. Hence, if the participant was instructed to decide whether each of these conclusions was necessary given the premise, the correct answer would be yes for 2 and 6 and no for all the others. If the instruction were to decide which were possible, then the correct answer would be yes for 2, 4, 6, and 7 and no for 1, 3, and 5.

In Experiment 1, participants were presented with immediate inference problems under either Necessity or Possibility instructions. This allows us to test predictions P1-P3 identified earlier. Note that in this design there is an inescapable confounding between the type of instruction (Possibility or Necessity) and the kind of answer constituting a correct response. For example, because many more conclusions are in fact possible than necessary, any general bias to accept conclusions would lead to more correct responding in the Possibility group compared with the Necessity group. However, this does not present us with a problem in testing our predictions above, which are in any case phrased in terms of conclusions accepted rather than correct decisions. Note also that predictions P2 and P3 both involve comparisons within groups, in which instruction type is held constant.

Method

Design. Participants were run in two experimental groups according to whether they received Necessity or Possibility instructions. Each evaluated the 28 immediate inference problems described above in an independently generated random order. Experiment 1 was the first task administered to these participants, who then proceeded to carry out the tasks of Experiment 2, reported below.

Participants. A total of 120 undergraduate students at the University of Plymouth, Plymouth, England, took part in this study and received payment for their participation. Participants were randomly allocated to the two experimental groups, 60 in each. All were familiar with the use of computers including the operation of a mouse.

Materials. The instructions for the tasks were presented separately on sheets of A4 paper. However, the reasoning problems were presented on an Acorn computer using a custom-written computer program. A separate screen was used for each of the 28

problems with a layout as in the example that follows:

GIVEN THAT
Some K are R
IS IT NECESSARY THAT
Some R are not K

The above example is in the form shown to the Necessity group. In the Possibility group, the second header read "IS IT POSSIBLE THAT." Underneath were two boxes showing the words YES and NO in uppercase. Participants signaled their decision by moving a mouse pointer into either box and clicking. The program randomly assigned letters to the sentences excluding I and O.

Procedure. Participants were run in small groups in a laboratory containing several computers running identical software. Each had their own workstation and worked independently at their own pace. Participants were randomly assigned to the experimental conditions on entering the room. As a group, they were given a brief introduction as to what the tasks of both Experiments 1 and 2 would involve, and they were informed that their data would be confidential and that they had a right to withdraw at any time without penalty.

The sheets containing the instructions for Experiment 1 were already present at the participants' computer terminals; they were asked to read these and then to start as soon as they were ready. They were instructed that the experiment was designed to find out how people solve logical reasoning problems. (Because naïve reasoners are known to assume that assertions of the form "All A are B" implicate the existence of A's and B's, we did not feel it necessary to state this explicitly in the instructions. Our results support our assumption.) It was explained that the problems would be presented one at a time on separate screens. The problem layouts were described, and participants were instructed to click in the YES or NO box to make their response. Necessity participants were instructed to click YES if the conclusion must follow from the premises and NO if it did not follow. Possibility participants were instructed to click YES if the conclusion could follow from the premises and NO if it could not. Each participant received the 28 problems in an independently randomized order. Both answers and response latencies¹ were recorded by the program and saved to a disk.

Results and Discussion

This study involves collection of data sets that could be analyzed for a number of theoretical purposes other than

¹ Although latency measures were collected in these experiments, we decided not to report them. Interpretation of the latency data was considered problematic for two main reasons. First, in the large experiment (Experiment 2) people had to evaluate four conclusions in relation to the premises pair on each screen with only one latency recorded. Hence, we cannot separate the time taken for reasoning about each conclusion. Second, there is much variability of responding (our main measure of interest) associated with different types of logical problems in all of the experiments, so that overall latencies are based on a variable mix of yes and no responses. There is no possibility of excluding "error" trials as with a simple cognitive task. Given the multifactorial nature of the designs, there is also no practical method of analyzing responses to yes and no decisions separately. In addition to these main problems, we are also of the view that psychologists currently lack a good theory of how response latencies map onto cognitive processes with problems of this complexity.

those that we have space to describe here. Consequently, a complete list of the percentage of conclusions endorsed on each problem by each group is shown in Appendix A. We concentrate our analysis here on the three categories of problems identified a priori as having conclusions that were Necessary ($n = 6$), Possible ($n = 12$), or Impossible ($n = 10$) given their premises. The mean percentage acceptance of conclusions in each category within the two instruction groups is shown in Figure 1. Note that more conclusions were endorsed in the Possibility group throughout and that the order of endorsement of problem type was Necessary > Possible > Impossible under both forms of instruction. The latter trend is to be expected, not just by the model theory but by any account that allows a significant element of deductive competence in syllogistic reasoning.

A 2×3 analysis of variance (ANOVA) was carried out to test the effect of groups (Necessity vs. Possibility) and types of inference (Necessary vs. Possible vs. Impossible). For purposes of this analysis, the percentage of inferences endorsed by each participant was computed for each problem type such that each participant contributed three data points to the ANOVA. There was a highly significant effect of group, reflecting the greater endorsement under Possibility instructions, $F(1, 118) = 50.89$, $MSE = 0.421$, $p < .001$. There was also a highly significant main effect of problem type, $F(2, 236) = 283.8$, $MSE = 0.349$, $p < .001$. In addition, the interaction between the two variables was significant, $F(2, 236) = 7.07$, $MSE = 0.349$, $p < .002$. The main effect of instructions confirms our general prediction P1, discussed earlier. The interaction appears to reflect the larger difference between the two groups for conclusions that were possible. The phenomenon is not surprising,

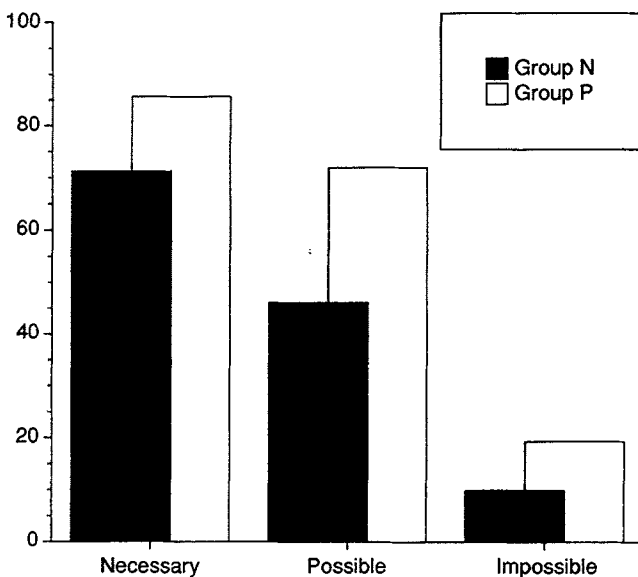


Figure 1. Percentage of conclusions accepted in Experiment 1 (immediate inference) in Group N (Necessity instructions) and Group P (Possibility instructions).

because these conclusions differ in their correct answer between the two groups: it is yes for the Possible group and no for the Necessary group.

Follow-up analyses were conducted to test predictions P2 and P3 derived from the mental model theory. As a reminder, P2 states:

P2. It will be easier to decide that a conclusion is possible if it is also necessary. Specifically, we predict more endorsements of possibility for Necessary than for Possible problems.

The rationale for this prediction is that people only need to identify one model that supports the conclusion to decide that it is possible. On Necessary problems all models of the premises support the conclusion, whereas on Possible problems there is at least one model that does not. Hence, it will be easier to find a supporting model on Necessary problems. This prediction was tested using a one-tailed related-groups t test comparing endorsement of Possible and Necessary problems for the group given Possibility instructions. The prediction was strongly confirmed, $t(59) = 4.84$, $p < .001$.

Prediction P3 states:

P3. It will be easier to decide that a conclusion is not necessary if it is also not possible. Specifically, we predict that more Possible than Impossible problems will be endorsed as necessary.

The rationale here is that reasoners need only identify one model of the premises that does *not* contain the conclusion in order to decide that the conclusion is not necessary. This will be easier on Impossible problems because no model of the premises supports the conclusion than on Possible problems in which at least one model will support it. This prediction was tested with a one-tailed t test comparing acceptance rates for Possible and Impossible problems under Necessity instructions and was again supported, $t(59) = 17.04$, $p < .001$.

All of the main predictions have then been strongly supported in our analysis of conclusions accepted on the immediate inference task of Experiment 1. An additional aspect of the frequency data is also worthy of a note. Whereas Possible problems are less frequently endorsed than Necessary problems overall in both instruction groups, the data under Necessity instructions seem to follow a bimodal distribution. Of the 12 Possible problems, the 7 most frequently endorsed ranged from 41% to 91% and the remaining 5 from 3% to 11%. The first group averages 75%—slightly more than the mean acceptance for Necessary problems with these instructions; the second group averages 6.5%—a little less than the mean for Impossible problems. Although retrospective, this observation suggests that reasoners often fail to search for counterexamples even when instructed to make decisions of necessity. The data are consistent with the hypothesis that some Possible problems suggest an initial model that supports the conclusion—leading to endorsement of the fallacy—whereas others do not. We return to this issue in discussing the results of Experiment 2 and providing the rationale for Experiment 3, both involving syllogistic reasoning proper.

Experiment 2: Syllogistic Reasoning

Experiment 2 extends the investigation reported in Experiment 1 to syllogistic inference, so we first provide some brief discussion of the logic and psychology of syllogisms. A syllogism has two end terms, which we refer to as A and C—in order of mention in the premises—and a middle or linking term which we call B. The conclusion can be in either order A–C or C–A. Either premise or the conclusion may take each of the four forms (moods) discussed earlier and investigated in the immediate inference task of Experiment 1. For example, a possible syllogism is

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

In classical syllogistic logic, a distinction is drawn between four figures of syllogism, which are defined by the order of the terms in the premises with reference to the order in the conclusion. Classical syllogisms are also defined in a way that fixes only one order of premises. In this study we wish to consider all possible ways of presenting these problems with conclusions. Hence, we follow the practice of Johnson-Laird and Bara (1984) in defining figure by the order of terms in the premises without regard to the conclusion. There are four such possible orderings:

- Figure 1 A–B
 B–C
- Figure 2 B–A
 C–B
- Figure 3 A–B
 C–B
- Figure 4 B–A
 B–C

In an evaluation task, as used in the present study, each of the above figures can be presented with a conclusion in either AC or CA order. The number of possible problems increases considerably in the evaluation paradigms because the two possible conclusion orders must be multiplied by the four possible moods. This results in 512 distinct problems. All of these were presented for evaluation in Experiment 2. To keep the task within reasonable bounds, however, participants evaluated either the AC or CA order but not both. In addition, they evaluated four conclusions on the same screen for a given premise pair, so that they effectively evaluated 256 syllogisms each, presented on 64 separate screens.

There are numerous theories of syllogistic inference to be found in the psychological literature (see Evans et al., 1993). The details of most of these need not concern us here. We note briefly that the theories fall into two broad types. One tradition emphasizes nonlogical processes in which the form of the premises biases the choice of conclusion without intervention of deductive processes. Although this approach has been out of fashion for many years, a new nonlogical theory based on probability heuristics has been developed recently by Chater and Oaksford (1999). With regard to deductive approaches, a provisional account of reasoning with quantifiers has been offered by Rips (1994) within the

mental logic tradition. A detailed account of syllogistic reasoning has been developed over the years by mental model theorists (see Johnson-Laird & Byrne, 1991) and has been implemented in a working computer program (Johnson-Laird & Byrne, 1994). A significant development is the verbal reasoning model of Polk and Newell (1995), which is also model based but differs in some significant respects from the theory presented by Johnson-Laird and his colleagues.

As stated earlier, our objective in this research is to test predictions derived from general principles of the model theory that were independent of implementation details. For this reason, we omit a description of the detailed account of syllogistic reasoning offered by Johnson-Laird and Byrne (1991), which is not required to follow the rationale and analysis of the present study. The experiment permits us to test the same three predictions P1, P2, and P3 about the difference between necessary and possible inference that were assessed on the simpler immediate inference task in Experiment 1.

Method

Design. Participants were run in four experimental groups immediately following the presentation of the immediate inference task reported as Experiment 1. Participants received the same kind of experimental instruction (Necessity or Possibility) as they were given in Experiment 1 but were subdivided according to whether they received conclusions to evaluate in AC or CA order (see introduction to this experiment). Each participant was shown 64 separate computer screens, one for each possible premise pair. For each pair they were required to indicate whether each of four possible conclusions was either necessary or possible according to instruction group. The four possible conclusions were those in each of the four moods and in an order (AC or CA) dependent on the subgroup to which participants were assigned.

Participants. The same 120 students who participated in Experiment 1 served as participants for Experiment 2, administered immediately afterward.

Materials. Problems were presented in a similar form to those of Experiment 1, except that there were two premises and a list of four conclusions to be evaluated. An example screen layout is as follows:

GIVEN THAT		
Some T are P		
No P are G		
IS IT NECESSARY THAT		
All T are G	YES	NO
Some T are G	<input type="checkbox"/>	<input type="checkbox"/>
No T are G	<input type="checkbox"/>	<input type="checkbox"/>
Some T are not G	<input type="checkbox"/>	<input type="checkbox"/>

The above example would be given to the Necessity-AC group. Possibility participants were shown a second header that read, IS IT POSSIBLE THAT, and those in CA groups would have received the conclusions reversed: All G are T, and so on. Letters were assigned randomly and independently to logical forms by the program on each problem. The order of the four conclusion types (moods) was fixed as above as A, I, E, O for all problems.

Procedure. On completion of Experiment 1, participants were presented with further written instructions titled TASK 2. They

were told that there would be another series of logical problems but that this time there will be two statements and that they would be asked, given that these statements were true, to decide which of the following statements must be true (Necessity group) or could be true (Possibility group). They were shown an example of the screen layout and were told to signal their answers by clicking the mouse in the YES or NO box next to each of the following statements. They were told that they could decide about the statements in any order but that a response once made could not be changed.

Whenever participants clicked in a box, it was filled on the screen. When all four responses were made to a screen, a box appeared underneath labeled **CLICK HERE FOR THE NEXT PROBLEM**. Thus participants were self-paced and could rest between problems in the long sequence as required. The program recorded both the responses and the latencies and saved them to disk. The 64 screens were presented in an independently randomized order for each participant.

Results and Discussion

A list of response frequencies to all individual problems is given in Appendix B. Of the 256 premise-conclusion combinations evaluated by each participant, 24 were Necessary (conclusion must follow from premise), 208 were Possible (conclusion could follow from premises), and 24 were Impossible (conclusion could not be true given the premises). These numbers reflect the fact that where, for example, premises support a conclusion of the form "All G are T," they also support "Some G are T" and "Some T are G," given that naive reasoners take universal assertions to imply the existence of the relevant entities and that "some" is taken to mean "at least some, and possibly all." The mean acceptance rates for each problem type in each instruction group are shown in Figure 2 averaged over the AC and CA subgroups (this factor had no overall effect). The graph is extraordinarily similar to that for Experiment 1 (see Figure

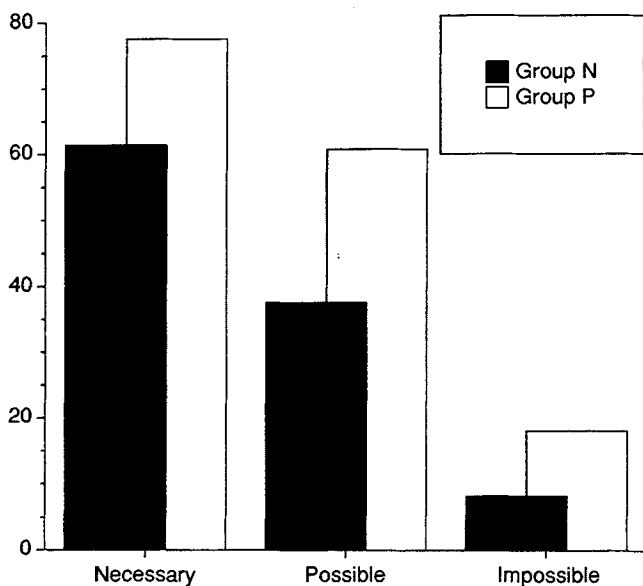


Figure 2. Percentage of conclusions accepted in Experiment 2 (syllogistic inference) in Group N (Necessity instructions) and Group P (Possibility instructions).

1) when one considers the additional complexity of the two premise syllogistic format. A $2 \times 2 \times 3$ ANOVA was conducted again on aggregated data with each individual contributing one data point for each of the three types of problems. The additional (group) factor to that of Experiment 1 was conclusion order.

The trends in the ANOVA were similar to those observed in Experiment 1. There was a significant main effect of instruction group, $F(1, 116) = 64.7$, $MSE = 0.374$, $p < .001$, reflecting higher acceptance rates under Possibility instructions. There was also a significant main effect of problem type, $F(2, 232) = 446.6$, $MSE = 0.218$, $p < .001$, reflecting the order Necessary > Possible > Impossible. Again, as in Experiment 1, the interaction between these two factors was significant, $F(2, 232) = 6.03$, $MSE = 0.218$, $p < .01$, reflecting the fact that instructions had the largest effect on problems of Possible type. There was, however, no effect of conclusion order.

As in Experiment 1, prediction P1 is confirmed as a main effect in the ANOVA, but P2 and P3 require follow-up tests. These were assessed using one-tailed related-groups t tests, with separate tests being carried out for the AC and CA subgroups ($dfs = 29$). P2 predicts that participants given Possibility instructions will accept more Necessary than Possible conclusions. This was confirmed for both AC ($t = 6.11$, $p < .001$) and CA ($t = 5.19$, $p < .001$) subgroups. P3 predicts that participants given Necessity instructions will accept more Possible than Impossible conclusions. This also was confirmed for both AC ($t = 12.3$, $p < .001$) and CA ($t = 9.2$, $p < .001$) subgroups.

This experiment provides the opportunity to look at some other issues of interest using the syllogistic evaluation task. An important claim of the model theory is that people will be less inclined to draw valid conclusions from syllogistic premises consistent with multiple mental models, as opposed to a single model. The point here is that when the conclusion is valid—as on our Necessary problems—then any model of the premises will support the conclusion. Hence, if people consider only one model of the premises, there should be no difference in performance on single and multimodel problems. If, on the other hand, people make an effort at deduction by trying to prove the conclusion in all models of the premises, then multimodel problems should be more difficult. To our knowledge, this prediction has not previously been tested on the syllogistic evaluation task, although evidence for it was adduced by Johnson-Laird and Bara (1984) using the production task. To maximize comparability with their study, we eliminated syllogisms with valid but weak conclusions from this analysis—for example, ones in which the conclusion to be evaluated was "Some A are C" even though the conclusion "All A are C" was supported by the premises. Such conclusions are rarely generated on a production task.

Of the 256 problems presented to each participant, just 36 are valid with strong conclusions. For this subset, we examined the difference in acceptance rates between single and multiple-model problems. Under Necessity instructions, we would expect a similar trend to that observed by Johnson-Laird and Bara (1984)—that is, fewer acceptances

of the valid conclusions on multiple-model problems. Under Possibility instructions, we should not find such a trend, however, because in this case only a single model need be found supporting the conclusion for the participant to justify a YES response. The trends in our data accord with this hypothesis. Of the 36 problems, 18 are of a kind classified as single model, and 18 as multiple model by Johnson-Laird and Bara. Under Necessity instructions, 81% of single model problems had their conclusions accepted, compared with 59% of multiple-model problems. As expected, this trend was weaker under Possibility instructions: 88% acceptance of single-model problems compared with 77% acceptance on multiple-model problems. To test the significance of this interaction, we computed the difference between acceptance rates of single- and multiple-model problems for all participants. These difference scores were then compared on a one-tailed between-groups *t* test, comparing the 60 participants who received Necessity instructions with the 60 in the Possibility group. The test was comfortably significant, $t(118) = 2.89, p < .005$.

The above finding supports the model theory claim that reasoners look for counterexamples, at least on this restricted set of 36 syllogisms. However, looking at the data as a whole, it is clear that any such facility must be weak, in view of the very high rates of endorsement of fallacies reported in this study as well as elsewhere in the literature. Nevertheless, it appears that the lower rate of acceptance of Possible than Necessary problems under Necessity instructions could be due to some searching for counterexamples that successfully leads to refutation of some of the potential fallacies. However, there is an alternative interpretation of the lower rate of acceptance of Possible conclusions that we need to consider. Suppose that people are willing to accept as necessary any conclusion supported by the *first* model of the premises they think of. Suppose also that the premises of Possible problems consistently suggest a model that either does or does not support the conclusion. The result would be that *some* Possible problems would be endorsed as frequently as Necessary problems and others as infrequently as Impossible problems. We have already noted that this was the case for the immediate inference task in Experiment 1. Inspection of the data for Possible problems under Necessity instructions in the syllogistic inference task of Experiment 2 suggested a similar trend was present. Some fallacies were strongly endorsed, as for example:

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

Ninety percent of participants judged the above conclusions to follow under instructions of Necessity, even though it is fallacious. Judgments of Possibility were also 90%, suggesting that (a) people easily imagine situations (mental model) in which premises and conclusion are true but (b) they fail to seek counterexamples to prove necessity. By contrast, consider the following syllogism:

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

Only 3% of participants endorsed this syllogism under instructions for Necessity, but 43% did so when given Possibility instructions. This suggests that participants do not initially think of a model of the premises that supports the conclusion in this case and hence reject it as unnecessary. However, in this case it appears that some effort to find an alternative and supporting model does occur when people are asked if the conclusion is possible. A trend to seek alternative models when proving possibility rather than necessity could account for the interaction between instructions and logical classification of problems reported above. Recall that this interaction reflects the fact that the increase in acceptance of conclusions under Possibility instructions was larger for Possible syllogisms than either Necessary or Impossible ones.

Experiment 3 was designed to establish further the reliability of the finding that some fallacies are consistently made and others withheld and to investigate the processes that underlie this effect.

Experiment 3: Possible Strong Versus Possible Weak Syllogisms

As noted earlier, Experiments 1 and 2 have strongly supported the a priori predictions that we derived about possible and necessary inference from the mental models account. However, there is also an observation that fits less comfortably with this framework. It appears that when Possible problems are assessed under Necessity instructions, some are accepted with very high rates comparable with those of Necessary problems, whereas others have very low acceptance rates comparable with those associated with Impossible problems (see above for examples of the two kinds of syllogism). A possible interpretation is that people are not searching for counterexample models, but rather are simply saying yes if the first model suggested by the premises contains the conclusion.

Experiment 2 is to our knowledge the only study to have examined people's evaluations to all possible conclusions with all possible premise types, and certainly the first to have done so with judgment of Possibility as well as Necessity required. Although it should prove useful to theorists of all persuasions to have such a database available (Appendix B), there is a disadvantage inherent in such a comprehensive approach, namely, that a very large number of problems have to be administered to each participant. This enforced the use of a method in which 64 premise pairs were presented with four conclusions to be evaluated, as the alternative would have required processing of 256 separate computer screens for each participant, even keeping conclusion order as a between-participant variable.²

² Another possible problem with the procedure of Experiment 2 is that the conclusions to be evaluated are logically related to each other. For example, if a participant has decided that the conclusion "All A are C" is necessary given the premises, then they might conclude that "Some A are not C" is impossible without returning

Table 1
Syllogisms Selected on the Basis of Inference Rates in Experiment 2 for Further Investigation in Experiment 3

Syllogism	A-B B-C		B-A C-B		A-B C-B		B-A C-B	
	ac	ca	ac	ca	ac	ca	ac	ca
Necessary	AA a	AE e	AI i	AA a	AE e	AE e	AI i	AA a
	AE e	IA i	AE o	AI i	IE o	AO o	AE o	AI i
	IA i	EA o	IE o	EA e	EA e	EA e	IE o	EA e
Impossible	IE o	EI o	EA e	EI o	OA o	EI o	EA e	EI o
	AE a	AA e	AA e	AA e	AE a	AE a	AA e	AA e
	AE i	AE a	AI e	AA o	AE i	AO a	AI e	AA o
Possible strong	IA e	AE i	IE a	EA a	EA a	EA a	EA a	EA a
	IE a	IA e	EA a	EI a	OA a	EI a	EA a	EI a
	AO o	AI i	AO o	II i	AI o	AI i	AO o	II i
Possible weak	II i	IO o	IA i	IO o	IA i	IA i	IA i	IO o
	OA o	OA o	II o	OA o	IO o	IO o	II o	OA o
	OO o	OI o	OA o	OI o	OI o	OA o	OA o	OI o
	II e	AO a	AI a	AO a	AI a	IA a	AI a	AO a
	OI e	II a	AO a	IE a	IA e	IE a	AO a	IE a
	OE a	IE a	OA a	EO a	II a	EO a	OA a	EO a
	OO a	OO a	OO a	OO a	OE a	OI a	OO a	OO a

Note. Uppercase letters represent premise pairs and lowercase represent conclusion. Key to premise and conclusion types: A = All X are Y; E = No X are Y; I = Some X are Y; O = Some X are not Y.

Experiment 3 involves presenting selected syllogisms such that each participant receives only 64 problems in total with each presented on a single screen, similar to the immediate inference problems of Experiment 1. In selecting syllogisms, our main interest was in the hypothesis that there are two kinds of Possible problems that we define as follows:

Possible Strong: problems whose conclusions are possible but not necessary given their premises, but that are frequently endorsed as having necessary conclusions.

Possible Weak: problems whose conclusions are possible but not necessary given their premises, and that are infrequently endorsed as having necessary conclusions.

In other words, Possible Strong syllogisms are the fallacies that people tend to make (under instructions for Necessity), and Possible Weak syllogisms are the fallacies they tend to avoid. The identification of these two categories has been necessarily retrospective in the two experiments reported to date. Hence, we decided to choose examples of Possible problems with high and low endorsement rates in Experiment 2 and submit them to independent replication in Experiment 3. The actual syllogisms selected are shown in Table 1. Note that the 64 problems presented to AC groups were not necessarily the same as those presented to CA groups. This is because problems were selected strictly on the basis of performance in Experiment 2 under Necessity instructions. The Necessary syllogisms selected were the four most frequently endorsed associated with each figure of premises and conclusion order. Possible Strong syllogisms were produced by choosing the four most frequently en-

dorsed Possible problems in each figure. Similarly, Possible Weak problems were the least frequently endorsed. Finally, the least frequently endorsed Impossible problems were selected. This method provides a strong test of whether Possible Strong problems are endorsed as often as Necessary problems, and Possible Weak problems as rarely as Impossible problems. With Necessity instructions, the levels were comparable in both cases (see the black bars in Figure 3). When the same problems were considered under Possibility instructions, however, Possible Weak problems had a somewhat higher acceptance rate than Impossible problems (see Figure 4). Although the Necessity data of Experiment 2 were used as the basis for problem selection, participants received the same syllogisms in Experiment 3 whether responding to Possibility or Necessity instructions.

Method

Design. There were four groups of participants (as with Experiment 2) according to conclusion order (AC or CA) and the type of instruction (Necessity or Possibility). Each participant was presented with 64 separate syllogisms generated by two within-participant variables on four levels each: the type of problem (Necessary, Impossible, Possible Strong, and Possible Weak), syllogistic figure, and a repetition of four examples in each cell.

Participants. Eighty undergraduate students from the University of Plymouth took part in this experiment and received either payment or course credits for their participation. None had participated in Experiment 1 and 2 or in other similar studies. They were randomly allocated to the four experimental conditions, in groups of 20.

Materials and procedure. The computer program randomly allocated letters (excluding I and O) for the end and middle terms used in the syllogisms that were previously selected as shown in Table 1. The screen layout used was similar to that of Experiment 1 except that two premises were given with a single conclusion to be

to the premises. Any possible distortion of responses resulting from this will be eliminated by the procedure in Experiment 3, in which only one conclusion is evaluated at a time with respect to a particular pair of premises.

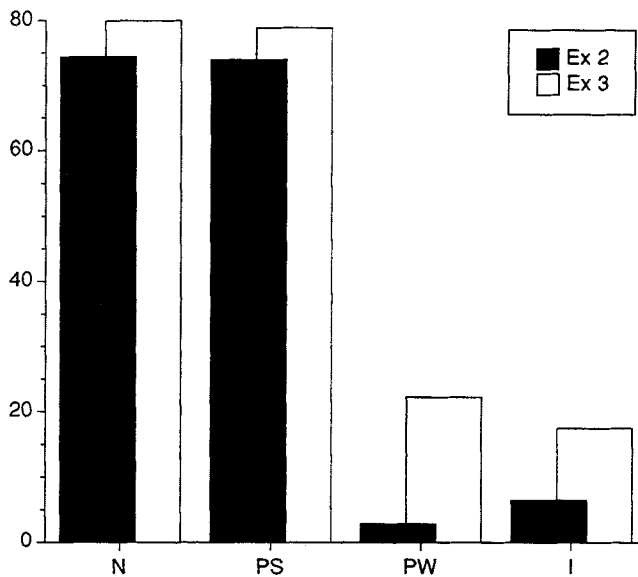


Figure 3. Percentage of conclusions accepted in Experiment 2 (Ex 2) under Necessity instructions broken down into four problem types: Necessary (N), Possible Strong (PS), Possible Weak (PW), and Impossible (I), together with replication data from Experiment 3 (Ex 3).

evaluated by pressing the mouse in a YES or NO box. The 64 syllogisms were presented on separate screens in an independently randomized order for each participant.

Initially the participants were given an overview of the structure of the experiment and informed of their right to withdraw and aspects of confidentiality. They were then presented with instruc-

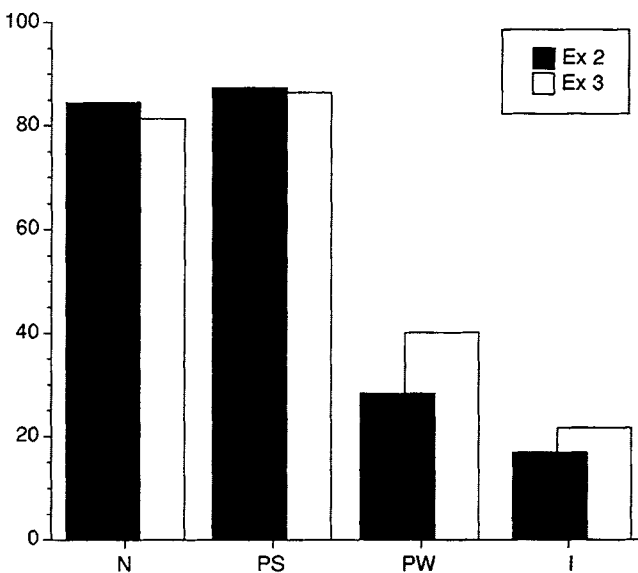


Figure 4. Percentage of conclusions accepted in Experiment 2 under Possibility instructions broken down into four problem types: Necessary (N), Possible Strong (PS), Possible Weak (PW), and Impossible (I), together with replication data from Experiment 3 (Ex 3).

tions on screen depending on the experimental condition to which they were assigned. Instructions were similar to those used in Experiment 2 except for the requirement to evaluate a single conclusion on each screen. Before the syllogisms were randomly presented, participants were given practice with the mouse to get used to the manner of signaling Yes and No responses.

Results and Discussion

The frequency of acceptance of different logical types under Necessity instructions is shown in Figure 3 together with the corresponding data from Experiment 2. It can be seen that Experiment 3 provides a close replication, dispelling any possibility that the difference between Possible Strong and Possible Weak problems was magnified by retrospective choice of these categories (e.g., due to statistical regression). It remains the case in Experiment 3 that acceptance rates of Necessary and Possible Strong problems are very similar. Possible Weak problems are also accepted only with slightly higher frequency than Impossible problems. Figure 4 compares acceptance frequencies for the different problem types under Possibility instructions, again comparing the data of Experiments 2 and 3. A good replication is again seen, but the tendency for Possible Weak problems to be more highly accepted than Impossible problems appears more marked in the replication.

Mean percentage acceptance of conclusions in Experiment 3 is shown in Table 2, broken down by figure, problem type, conclusion order, and instruction type. Each cell represents the mean of 20 participants' responses with four replications of syllogisms. The use of replications permitted us to run a four-way ANOVA on the frequency of Yes responses, with instruction and conclusion order as between-groups variables and figure and problem type as within-group variables. There was a significant main effect of instruction, $F(1, 76) = 18.51$, $MSE = 0.127$, $p < .001$, such that more conclusions were accepted under Possibility (57%) than Necessity (49%) instructions and as might be

Table 2
Percentage of Conclusions Accepted in Experiment 3,
Broken Down by Problem Type, Figure, Conclusion Order,
and Instructional Group

Instruction and problem type	A-B B-C		B-A C-B		A-B C-B		B-A B-C		<i>M</i>
	ac	ca	ac	ca	ac	ca	ac	ca	
Necessity									
Necessary	81	79	70	84	76	86	84	79	80
Impossible	9	21	25	20	13	15	24	13	17
Possible strong	84	83	80	81	74	81	70	78	79
Possible weak	18	23	24	19	19	15	13	19	19
<i>M</i>	48	51	50	51	46	49	48	47	
Possibility									
Necessary	89	78	86	86	74	83	86	85	83
Impossible	31	15	30	5	21	14	30	19	21
Possible strong	88	93	86	93	90	91	79	79	87
Possible weak	59	26	51	21	45	26	43	29	38
<i>M</i>	67	53	64	51	58	54	59	53	

expected a very highly significant effect of problem type, $F(3, 228) = 412.00$, $MSE = 0.090$, $p < .001$.

In Experiments 1 and 2, problem type (defined on three rather than four levels) interacted significantly with instructions, such that relatively more Possible conclusions were accepted under instructions for Possibility. The two variables also interacted significantly in this analysis, $F(3, 228) = 4.85$, $MSE = 0.903$, $p < .005$. With the breakdown of the Possible category into Strong and Weak, however, it appears from the means (right-hand column of Table 2) that the major difference lies in the acceptance rates of Possible Weak problems that are accepted at a rate of 38% under Possibility instructions compared with only 19% under Necessity instructions. This difference is statistically significant, $t(78) = 4.48$, $p < .01$, and is theoretically important because it suggests that under instructions for Possibility, reasoners *do* seek alternative models of the premises if the first fails to support the premises, 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.

There were several other significant effects. First, there was a significant if small main effect of figure, $F(3, 228) = 3.04$, $MSE = 0.026$, $p < .05$. Conclusions were most often accepted in Figure 1 (55%) and least often in Figure 4 (52%). There was no main effect of conclusion order, but this variable did interact significantly with instruction, $F(1, 76) = 7.88$, $MSE = 0.127$, $p < .01$. Inspection of Table 3 suggests that this was due to greater acceptance of AC conclusions in the Possibility group only. There is no obvious explanation for this trend. The remaining trends were interactions involving problem type. This interacted significantly with figure and conclusion order separately, and all three variables significantly if weakly interacted together, $F(9, 684) = 2.12$, $MSE = 2.12$, $p < .05$.

Finally, it should be noted that the strong replication of similar findings for the same syllogisms presented in Experiment 3 when compared with Experiment 2 counters two possible criticisms of the design of the previous experiment. First, it might be suggested that the prior experience of participating in Experiment 1 might have had some effect on performance in Experiment 2; Experiment 3 had no such prior task. Second, it could be argued that participants became fatigued or bored while processing the large number of syllogisms used in Experiment 2, which might add noise to their data. We think this was unlikely to be true in any case, because of the self-pacing provided and the fact that the analysis showed very clear and systematic trends. The fact that the data were so similar on corresponding syllogisms in Experiment 3, in which only one quarter as many were used, confirms this view.

General Discussion

Our experiments have shown that logically untrained individuals are able to draw modal conclusions both to single quantified assertions and to syllogistic premises. The

results also confirm our three main predictions based on the mental model theory. First, the participants were more likely to endorse conclusions about what is possible than what is necessary (Prediction P1). Possibility calls for only a single model of the premises to support the conclusion, whereas necessity calls for all the models of the premises to support the conclusion. Second, the participants were more likely to endorse a possible conclusion if the premises supported its necessity as opposed to its mere possibility (Prediction P2). In the case of necessity, the conclusion holds in all the models of the premises, and so no matter which model of the premises reasoners construct, it will support the conclusion. In contrast, if the premises only support a possible conclusion, then reasoners need to search for the model of the premise in which the conclusion holds, and they need to ignore those models that are counterexamples to the conclusion.

Our third hypothesis was that participants would be more likely to decide that a conclusion was not necessary if it was impossible (Prediction P3). A conclusion that is impossible holds in no models of the premises, whereas a conclusion that is merely not necessary holds in at least one model of the premises, and so reasoners are liable to go wrong in the latter case if they focus on a model in which the conclusion holds. In other words, they have to search for a counterexample to the content of the conclusion to infer that it is not necessary. The search is easier where all the models of the premises are counterexamples than where at least one is not. The three predictions were corroborated in Experiment 1 in which the participants evaluated immediate inferences from a single quantified premise to a quantified conclusion, in Experiment 2 in which they evaluated all possible modal conclusions from syllogistic premises, and in Experiment 3 in which they evaluated a subset of modal conclusions from syllogistic premises.

One initially unexpected effect was that problems supporting possible conclusions fell into two categories. Some of them were regularly taken to imply necessary conclusions, whereas others were rarely taken to imply necessary conclusions and indeed sometimes were not even taken to imply possible conclusions. This phenomenon was first detected in Experiment 1. It also occurred in Experiment 2, and it was corroborated by Experiment 3, which we designed to contrast the two sorts of problems, which we termed Possible Strong and Possible Weak. It is particularly striking in Figure 3 that Possible Strong syllogisms are endorsed as frequently as Necessary ones and that Possible Weak syllogisms are endorsed as infrequently as Impossible ones. This suggests that any search for counterexample models is weak in the present study and that most participants are basing their conclusions on the first model that occurs to them (we return to this issue later). On this argument, the fallacies are sharply divided into our two categories because some premise pairs consistently suggest an initial model that supports the conclusion (Possible Strong), whereas others consistently suggest a model that negates the conclusion (Possible Weak).

It is clear that such a finding cannot be accounted for at the

very general level of our other predictions. However, the computer program implementation of the model theory reported by Johnson-Laird and Byrne (1994) does generate models in specific orders, so we were able to examine the output of the program for our two categories of syllogisms to try to confirm our intuitive account. The results provided strong corroboration. All premise pairs studied in the two Possible conditions of Experiment 3 were run through the program. In 29 out of the 32 cases, the Possible Strong syllogisms had an initial model that supported the given conclusion, whereas none of the 32 Possible Weak syllogisms had an initial model that supported the given conclusion.

To see how this arises, consider an example of Possible Strong syllogisms whose fallacious conclusion is endorsed by most participants:

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

The initial model of the premises, as constructed by the program, is

a ¬b
a ¬b
 [b] c
 [b] c

where ¬ denotes negation and a, b, and c are tokens representing exemplars for the classes A, B, and C. The square brackets around b in the latter models indicate that b's are exhaustively represented with respect to c's. In other words, there cannot be b's found in models that do not contain c's. This model supports the conclusion presented, and 75% of the participants in Experiment 3 inferred (wrongly) that it was a necessary conclusion whereas 85% of the participants inferred (correctly) that it was a possible conclusion. In fact, as the program also shows, there is an alternative model of the premises that refutes the conclusion:

a ¬b c
a ¬b c
 [b] c
 [b] c

Hence, the conclusion is possible but not necessary. An example of a Possible Weak syllogism has the following form:

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

The initial model of the premises is:

[a] b c
[a] b
 c

This model does not support the given conclusion, but instead the conclusion "Some A are C" (or its converse). But this conclusion can also be refuted by an alternative model

of the premises:

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

To grasp that the given conclusion is at least possible, reasoners need to construct another model of the premises:

[a] b c
[a] b c
 c

As this account predicts, few participants in Experiment 3 wrongly inferred that the given conclusion was necessary (15%), and only a minority inferred that it was even possible (40%). It is important to note that discovery of such Possible Weak syllogisms was made in this experiment because of the methodology involved whereby participants were asked to evaluate every possible conclusion for every possible premise pair. With a production task method in which people are instructed to produce a single, necessary conclusion, the premise pairs involved would normally be associated with the production of some alternative conclusion, as illustrated in the above example. An interesting possibility for future research would involve asking people to produce conclusions that are possible rather than necessary.

The program implementing the model theory works by constructing separate models of the two premises and then joining them together by way of the tokens representing the middle term. Thus, additional steps are required to add further tokens to the model (as in the example above of a Possible Strong syllogism) or to split an individual, such as [a] b c, into two distinct individuals: [a] b and b c (as in the example above of a Possible Weak conclusion). These steps yield models that are constructed after the initial models (see Johnson-Laird and Byrne, 1994, for a detailed description of the operations of the program).

Although the above analysis is very encouraging for the model theory, we should note that alternative explanations can be offered for the difference between Possible Strong and Possible Weak syllogisms. In particular, the probability heuristics model of Chater and Oaksford (1999) can account for the difference in terms of the "min-heuristic." The min-heuristic specifies that the conclusion drawn from syllogistic premises will match the form of least informative premise, in which informativeness has the following order: all > some > no > some . . . not. Of the 32 Possible Strong syllogisms shown in Table 1, 30 have conclusions that match the less informative premise, whereas the remaining 2 have conclusions that are less informative than either premise. For the 32 Possible Weak problems, all but 1 have a conclusion that is more informative than at least one premise.

The equally strong account that both the model program and the min-heuristic provide the difference between Possible Strong and Weak syllogisms is both surprising and interesting. The min-heuristic is offered as part of an entirely nondeductive account of syllogistic reasoning by Chater and Oaksford (1999). The effect that it has is to ensure that arguments are not supported when their conclusions are

more informative than the premises. This property is true not only of all valid arguments (deduction cannot add information) but of many fallacies also. Hence, the theory accounts (as it was designed to do) for the fact that in syllogistic reasoning experiments, participants endorse most valid arguments but also many fallacies. It is clear that the model program, based on a quite different approach, has the effect of first generating models whose provisional conclusions conform with the min-heuristic. This coincidence has been discussed recently in terms of the “atmosphere effect”—whose predictions are similar to those of the min-heuristic—by Shaw and Johnson-Laird (1998). These authors also examined the order with which the mental models program produces models and observed, as we do here, that errors often result from fallacious conclusions being supported by the initial models of the premises considered.

Let us return now to the question of whether reasoners actually search for alternative models as envisaged by the third stage of the model theory. This issue is theoretically important because the emphasis placed on this stage by Polk and Newell (1995) in their verbal reasoning theory is much less than in the mental models account of, for example, Johnson-Laird and Byrne (1991). It is also the aspect of the model theory required to account for deductive competence. We say above that the evidence for such searching is weak in the present study. We do not say that it is *absent* because other aspects of our data—as well those of other studies in the literature—suggest that at least some search for alternative models is occurring. First, a robust result in standard syllogistic reasoning is that one-model problems are reliably easier than multiple-model problems—every study in the literature, including those carried out by critics of the model theory, has corroborated this result (see, e.g., Ford, 1995). This difference is manifest on valid syllogisms taken separately in which *any* model of the premises supports the conclusion. Thus if people were deciding on the basis of the first model alone, there should be no difference.

The present study provides some important new evidence for this claim. First, Experiment 2 showed that the effect occurs when participants were instructed to judge necessity of conclusions—equivalent to the validity judgments used in conventional reasoning experiments. Second, for those participants who were instructed to judge the possibility of conclusions, the effect was reliably reduced. This interaction strongly suggests that people understand that a judgment of necessity requires finding a conclusion in all models, whereas judgment of possibility does not. This requires the further assumption, however, that people realize that there are multiple models but are not confident of their ability to check them all.

Also pertinent is the significant interaction between instruction type and logical classification reported in Experiments 1 and 2. Inspection of Figures 1 and 2 reveals that this was a result of the difference in endorsement rates between Possibility and Necessity instructions that was most marked on Possible syllogisms (potential fallacies), which were not subdivided at this stage of our report. Although a general increase in acceptance rates under Possibility instructions could be due to response bias, such as a caution effect, the

interaction suggests that at least some people are searching for alternative models and finding them. However, the analysis of Experiment 3, in which the two kinds of Possible problems were separated, indicates that this difference was mostly due to a quite marked increase in acceptance rate of Possible Weak syllogisms under Possibility instructions. Thus we have clear evidence that people search for alternative models to prove the possibility of conclusions that are not supported by the first model considered, but in this case less clear evidence that people search for counterexamples to establish the necessity of conclusions that *are* supported by the first model considered.

The issue of whether people search for alternative models has also been investigated in studies reported elsewhere, with mixed conclusions depending on the methodology used. Newstead, Handley, and Buck (1999) used two process-tracing methods to address this question. The first method involved asking participants to indicate any other conclusion they had considered immediately after responding, and the second involved drawing Euler circle diagrams. Neither method produced evidence that people were considering alternative models. A rather different picture emerges from the study of Bucciarelli and Johnson-Laird (1998). They asked their participants to construct external models (cut-out shapes) as they drew conclusions from syllogistic premises, and they videotaped the resulting sequences of models. Their results showed that reasoners consider a variety of different interpretations of the premises, as claimed by Polk and Newell (1995). However, the participants tended not to reencode the premises for one-model syllogisms (in contrast with the verbal reasoning), and they did generate sequences of models for multiple-model syllogisms. Whether they were searching for counterexamples was less clear. However, another phenomenon of syllogistic reasoning does support a search for counterexamples. Byrne and Johnson-Laird (1990) tested their participants' ability to recognize conclusions that they had drawn earlier to syllogistic premises. As a search for counterexamples predicts, participants often falsely recognized conclusions supported by an initial model of the premises when, in fact, they had responded correctly that nothing followed from the premises. This result suggests that they had fleetingly considered the erroneous conclusion only to reject it as a result of a counterexample.

In view of this somewhat ambiguous evidence, a cautious conclusion might be that people *can* search for alternative models but do not necessarily do so spontaneously. This might explain why evidence for such search seems to depend on the experimental method used. We are aware of two factors that seem to influence strongly whether such a search occurs. The first is the so-called *belief bias* effect that might better be described as a debiasing effect. Although commonly described as a tendency for people to endorse fallacies whose conclusions are believable, it might be more accurately described as a tendency to suppress fallacies when their conclusions are unbelievable. Most problems used in this literature are of the type we term Possible Strong in which people will normally endorse the fallacy when the conclusion is abstract or neutral. Those few studies that have

included neutral as well as believable and unbelievable conclusions have reported that the endorsements of neutral conclusions are at least as high as for believable ones, and that belief bias is due to suppression of endorsement of unbelievable conclusions (Evans & Handley, 1997; Evans & Pollard, 1990; Newstead, Pollard, Evans, & Allen, 1992).

This finding supports the mental model account of the belief bias effect (see, e.g., Oakhill, Johnson-Laird, & Garnham, 1989). When a putative conclusion is unbelievable, then people are motivated to search for counterexamples to disprove it. Also consistent with the view that people may require some stimulus to engage in model searching is the effect of instructions. Evans, Allen, Newstead, and Pollard (1994) showed that strong instructional emphasis on logical necessity significantly reduced belief bias and also significantly suppressed the tendency to endorse possible but not necessary conclusions. If the natural mode of thinking is inductive rather than deductive (as argued by Evans & Over, 1996) then it makes sense that deductive competence will be most apparent under conditions in which people are clearly motivated to make an effort at deduction.

Finally, might there be an alternative explanation for our main findings in terms of a theory based on formal rules? Our principal predictions are independent of the details of the model theory, and they cannot be made by current versions of formal rule theories that describe only a mechanism for necessary inference. Osherson's (1976) rules for modal inferences concern only a limited set of inferences based on propositional connectives and make no predictions about quantified modal reasoning. Our findings clearly provide a strong challenge for mental logicians as well as those psychologists who believe that syllogistic reasoning can be explained without any reference to an effort at deduction.

References

- Bell, V. A., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science*, 22, 25–52.
- Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, 85, 1–21.
- Braine, M. D. S., & O'Brien, D. P. (1991). A theory of If: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, 98, 182–203.
- Braine, M. D. S., Reiser, B. J., & Rumin, B. (1984). Some empirical justification for a theory of natural propositional logic. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 18, pp. 317–371). New York: Academic Press.
- Bucciarelli, M., & Johnson-Laird, P. N. (1998). *Strategies in syllogistic reasoning*. Unpublished manuscript, Princeton University.
- Byrne, R. M. J., & Johnson-Laird, P. N. (1990). Remembering conclusions we have inferred: What biases reveal. In J.-P. Caverni, J.-M. Fabre, & M. Gonzalez (Eds.), *Advances in psychology: Vol. 68. Cognitive biases* (pp. 109–120). Amsterdam: North-Holland.
- Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology*, 38, 191–258.
- Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*. Hillsdale, NJ: Erlbaum.
- Evans, J. St. B. T., Allen, J. L., Newstead, S. E., & Pollard, P. (1994). Debiasing by instruction: The case of belief bias. *European Journal of Cognitive Psychology*, 6, 263–285.
- Evans, J. St. B. T., & Handley, S. J. (1997). *Necessary and possible inferences: A test of the mental model theory of reasoning* (Report to the Economic and Social Research Council, Grant R000221742). Plymouth, United Kingdom, University of Plymouth.
- Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove, England: Erlbaum.
- Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, England: Psychology Press.
- Evans, J. St. B. T., & Over, D. E. (1997). Rationality in reasoning: The case of deductive competence. *Current Psychology of Cognition*, 16, 3–38.
- Evans, J. St. B. T., & Pollard, P. (1990). Belief bias and problem complexity in deductive reasoning. In J. P. Caverni, J. M. Fabre, & M. Gonzalez (Eds.), *Cognitive biases* (pp. 131–154). Amsterdam: North-Holland.
- Ford, M. (1995). Two modes of mental representation and problem solution in syllogistic reasoning. *Journal of Experimental Psychology: General*, 115, 16–25.
- Galotti, K. M., Baron, J., & Sabini, J. P. (1986). Individual differences in syllogistic reasoning: Deduction rules or mental models? *Journal of Experimental Psychology: General*, 115, 16–25.
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, England: Cambridge University Press.
- Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition*, 16, 1–62.
- Johnson-Laird, P. N., & Byrne, R. (1991). *Deduction*. Hove, England: Erlbaum.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1994). Models, necessity and the search for counter-examples. A reply to Martin-Cordero & Gonzalez-Labra and to Smith. *Behavioral and Brain Sciences*, 17, 775–777.
- Manktelow, K. I., & Over, D. (1993). (Eds.). *Rationality*. London: Routledge.
- Newstead, S. E., & Griggs, R. A. (1983). Drawing inferences from quantified statements: A study of the square of opposition. *Journal of Verbal Learning and Verbal Behavior*, 22, 535–546.
- Newstead, S. E., Handley, S. J., & Buck, E. (1999). Falsifying mental models: Testing the predictions of theories of syllogistic reasoning. *Journal of Memory and Language*, 27, 344–354.
- Newstead, S. E., Pollard, P., Evans, J. St. B. T., & Allen, J. (1992). The source of belief bias in syllogistic reasoning. *Cognition*, 45, 257–284.
- Oakhill, J., Johnson-Laird, P. N., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition*, 31, 117–140.
- Osherson, D. N. (1976). *Logical abilities in children: Vol. 4. Reasoning and concepts*. Hillsdale, NJ: Erlbaum.
- Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review*, 102, 533–566.
- Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review*, 90, 38–71.
- Rips, L. J. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.
- Shaw, V., & Johnson-Laird, P. N. (1998). Dispelling the “atmosphere” effect on reasoning. In A. C. Quelhas (Ed.), *Cognition and context* (pp. 169–200). Lisbon: Instituto Superior de Psicologia Aplicada.

Appendix A

Percentage Endorsements of Conclusions in Experiment 1

Premise	Conclusion	Logic	Necessity	Possibility
All A are B	Some A are B	N	65	82
	No A are B	I	3	12
	Some A are not B	I	10	7
	All B are A	P	57	85
	Some B are A	N	72	80
	No B are A	I	17	32
	Some B are not A	P	42	52
Some A are B	All A are B	P	3	53
	No A are B	I	8	55
	Some A are not B	P	92	97
	All B are A	P	5	53
	Some B are A	N	82	95
	No B are A	I	22	50
	Some B are not A	P	85	97
No A are B	All A are B	I	3	2
	Some A are B	I	7	5
	Some A are not B	N	77	80
	All B are A	I	7	25
	Some B are A	I	20	35
	No B are A	N	67	90
	Some B are not A	N	65	87
Some A are not B	All A are B	I	2	3
	Some A are B	P	87	97
	No A are B	P	5	48
	All B are A	P	7	38
	Some B are A	P	70	95
	No B are A	P	12	53
	Some B are not A	P	88	95

Note. N = necessary; I = impossible; P = possible.

(Appendix B follows)

Appendix B

Percentage Endorsement of Conclusions in Experiment 2 Under Necessity (N) and Possibility (P) Instructions

Fig.	Conc. order	Premises	Conclusion	L	N	P	Fig.	Conc. order	Premises	Conclusion	L	N	P
1	A-C	All A are B All B are C	All A are C Some A are C No A are C Some A are not C	N	73	80	1	A-C	All A are B No B are C	All A are C Some A are C No A are C Some A are not C	I	3	10
1	C-A	All A are B All B are C	All C are A Some C are A No C are A Some C are not A	P	77	83	1	C-A	All A are B No B are C	All C are A Some C are A No C are A Some C are not A	I	7	13
1	A-C	All A are B Some B are C	All A are C Some A are C No A are C Some A are not C	P	10	30	1	A-C	All A are B Some B are not C	All A are C Some A are C No A are C Some A are not C	P	10	10
1	C-A	All A are B Some B are C	All C are A Some C are A No C are A Some C are not A	P	10	33	1	C-A	All A are B Some B are not C	All C are A Some C are A No C are A Some C are not A	P	6	27
1	A-C	Some A are B All B are C	All A are C Some A are C No A are C Some A are not C	P	7	23	1	A-C	Some A are B No B are C	All A are C Some A are C No A are C Some A are not C	I	0	13
1	C-A	Some A are B All B are C	All C are A Some C are A No C are A Some C are not A	P	13	37	1	C-A	Some A are B No B are C	All C are A Some C are A No C are A Some C are not A	P	3	20
1	A-C	Some A are B Some B are C	All A are C Some A are C No A are C Some A are not C	P	13	33	1	A-C	Some A are B Some B are not C	All A are C Some A are C No A are C Some A are not C	P	13	37
1	C-A	Some A are B Some B are C	All C are A Some C are A No C are A Some C are not A	P	3	50	1	C-A	Some A are B Some B are not C	All C are A Some C are A No C are A Some C are not C	P	7	53
1	A-C	No A are B All B are C	All A are C Some A are C No A are C Some A are not C	P	10	20	1	A-C	No A are B No B are C	All A are C Some A are C No A are C Some A are not C	P	17	30
1	C-A	No A are B All B are C	All C are A Some C are A No C are A Some C are not A	P	13	23	1	C-A	No A are B No B are C	All C are A Some C are A No C are A Some C are not A	P	17	37
1	A-C	No A are B Some B are C	All A are C Some A are C No A are C Some A are not C	P	13	27	1	A-C	No A are B Some B are not C	All A are C Some A are C No A are C Some A are not C	P	13	43
1	C-A	No A are B Some B are C	All C are A Some C are A No C are A Some C are not A	P	13	20	1	C-A	No A are B Some B are not C	All C are A Some C are A No C are A Some C are not C	P	17	30
1	A-C	Some A are not B All B are C	All A are C Some A are C No A are C Some A are not C	P	7	27	1	A-C	Some A are not B No B are C	All A are C Some A are C No A are C Some A are not C	P	3	23
1	C-A	Some A are not B All B are C	All C are A Some C are A No C are A Some C are not A	P	7	10	1	C-A	Some A are not B No B are C	All C are A Some C are A No C are A Some C are not A	P	10	37
1	A-C	Some A are not B Some B are C	All A are C Some A are C No A are C Some A are not C	P	13	20	1	A-C	Some A are not B Some B are not C	All A are C Some A are C No A are C Some A are not C	P	7	20
1	C-A	Some A are not B Some B are C	All C are A Some C are A No C are A Some C are not A	P	7	33	1	C-A	Some A are not B Some B are not C	All C are A Some C are A No C are A Some C are not A	P	0	23

Appendix B (continued)

Fig.	Conc. order	Premises	Conclusion	L	N	P	Fig.	Conc. order	Premises	Conclusion	L	N	P
2	A-C	All B are A All C are B	All A are C Some A are C No A are C Some A are not C	P N I P	63 43 3 33	83 80 7 30	2	A-C	All B are A No C are B	All A are C Some A are C No A are C Some A are not C	I P P N	17 17 23 27	20 30 80 67
2	C-A	All B are A All C are B	All C are A Some C are A No C are A Some C are not A	N N I I	80 33 10 7	87 70 17 20	2	C-A	All B are A No C are B	All C are A Some C are A No C are A Some A are not C	P P P P	10 20 70 47	27 37 80 60
2	A-C	All B are A Some C are B	All A are C Some A are C No A are C Some A are not C	P N I P	3 87 7 73	33 87 27 87	2	A-C	All B are A Some C are not B	All A are C Some A are C No A are C Some A are not C	P P P P	7 70 10 93	17 93 53 93
2	C-A	All B are A Some C are B	All C are A Some C are A No C are A Some C are not A	P N I P	10 83 13 53	27 87 20 80	2	C-A	All A are B Some B are not C	All C are A Some C are A No C are A Some C are not A	P P P P	3 53 20 70	17 87 23 90
2	A-C	Some B are A All C are B	All A are C Some A are C No A are C Some A are not C	P P P P	17 83 10 50	37 97 20 87	2	A-C	Some B are A No C are B	All A are C Some A are C No A are C Some A are not C	I P P N	10 27 57 50	20 47 83 83
2	C-A	Some B are A All C are B	All C are A Some C are A No C are A Some C are not A	P P P P	10 77 7 63	43 77 20 77	2	C-A	Some B are A No C are B	All C are A Some C are A No C are A Some C are not A	P P P P	0 20 60 67	27 63 70 77
2	A-C	Some B are A Some C are B	All A are C Some A are C No A are C Some A are not C	P P P P	3 77 7 83	53 97 57 87	2	A-C	Some B are A Some C are not B	All A are C Some A are C No A are C Some A are not C	P P P P	10 60 7 73	30 100 47 93
2	C-A	Some B are A Some C are B	All C are A Some C are A No C are A Some C are not A	P P P P	13 83 3 70	40 100 57 97	2	C-A	Some B are A Some C are not B	All C are A Some C are A No C are A Some C are not A	P P P P	3 57 7 77	33 87 43 93
2	A-C	No B are A All C are B	All A are C Some A are C No A are C Some A are not C	I I N N	7 13 73 33	17 20 97 70	2	A-C	No B are A No C are B	All A are C Some A are C No A are C Some A are not C	P P P P	17 27 53 33	30 30 93 67
2	C-A	No B are A All C are B	All C are A Some C are A No C are A Some C are not A	I I N N	10 10 77 30	13 17 83 60	2	C-A	No B are A No C are B	All C are A Some C are A No C are A Some C are not A	P P P P	23 23 53 37	37 47 83 63
2	A-C	No B are A Some C are B	All A are C Some A are C No A are C Some A are not C	P P P P	10 40 40 53	27 50 77 60	2	A-C	No B are A Some C are not B	All A are C Some A are C No A are C Some A are not C	P P P P	10 53 27 53	40 80 67 77
2	C-A	No B are A Some C are B	All C are A Some C are A No C are A Some C are not A	I P P N	3 43 33 73	10 63 70 80	2	C-A	No B are A Some C are not B	All C are A Some C are A No C are A Some C are not A	P P P P	0 47 27 53	17 53 83 77
2	A-C	Some B are not A All C are B	All A are C Some A are C No A are C Some A are not C	P P P P	7 70 10 87	27 87 53 77	2	A-C	Some B are not A No C are B	All A are C Some A are C No A are C Some A are not C	P P P P	10 47 30 53	27 60 73 70
2	C-A	Some B are not A All C are B	All C are A Some C are A No C are A Some C are not A	P P P P	3 57 13 80	23 83 27 97	2	C-A	Some B are not A No C are B	All C are A Some C are A No C are A Some C are not A	P P P P	10 30 37 43	23 53 63 57
2	A-C	Some B are not A Some C are B	All A are C Some A are C No A are C Some A are not C	P P P P	7 63 10 73	37 87 47 90	2	A-C	Some B are not A No C are B	All A are C Some A are C No A are C Some A are not C	P P P P	3 80 13 80	27 90 43 90
2	C-A	Some B are not A Some C are B	All C are A Some C are A No C are A Some C are not A	P P P P	10 57 17 80	30 90 43 87	2	C-A	Some B are not A Some C are not B	All C are A Some C are A No C are A Some C are not A	P P P P	0 67 7 77	20 97 50 87

Appendix B (continued)

Fig.	Conc. order	Premises	Conclusion	L	N	P	Fig.	Conc. order	Premises	Conclusion	L	N	P
3	A-C	All A are B All C are B	All A are C Some A are C No A are C Some A are not C	P P P P	60 27 20 23	83 63 23 30	3	A-C	All A are B No C are B	All A are C Some A are C No A are C Some A are not C	I I N N	3 7 83 33	7 20 87 63
3	C-A	All A are B All C are B	All C are A Some C are A No C are A Some C are not A	P P P P	67 47 17 17	77 60 23 30	3	C-A	All A are B No C are B	All C are A Some C are A No C are A Some C are not A	I I N N	7 7 80 37	13 17 80 60
3	A-C	All A are B Some C are B	All A are C Some A are C No A are C Some A are not C	P P P P	3 67 23 80	30 70 53 93	3	A-C	All A are B Some C are not B	All A are C Some A are C No A are C Some A are not C	P P P P	10 67 13 80	17 80 47 87
3	C-A	All A are B Some C are B	All C are A Some C are A No C are A Some C are not A	P P P P	10 80 23 70	23 90 33 80	3	C-A	All A are B Some C are not B	All C are A Some C are A No C are A Some C are not A	I P P N	0 57 17 80	10 73 43 83
3	A-C	Some A are B All C are B	All A are C Some A are C No A are C Some A are not C	P P P P	10 83 0 67	30 97 37 83	3	A-C	Some A are B No C are B	All A are C Some A are C No A are C Some A are not C	I P P N	13 37 33 63	17 63 80 77
3	C-A	Some A are B All C are B	All C are A Some C are A No C are A Some C are not A	P P P P	3 70 13 57	43 90 30 87	3	C-A	Some A are B No C are B	All C are A Some C are A No C are A Some C are not A	P P P P	0 37 43 40	17 53 80 73
3	A-C	Some A are B Some C are B	All A are C Some A are C No A are C Some A are not C	P P P P	3 57 10 70	40 90 47 87	3	A-C	Some A are B Some C are not B	All A are C Some A are C No A are C Some A are not C	P P P P	10 63 13 80	23 97 57 90
3	C-A	Some A are B Some C are B	All C are A Some C are A No C are A Some C are not A	P P P P	10 63 7 63	33 93 53 93	3	C-A	Some A are B Some C are not B	All C are A Some C are A No C are A Some C are not A	P P P P	7 63 7 73	43 87 57 83
3	A-C	No A are B All C are B	All A are C Some A are C No A are C Some A are not C	I I N N	10 17 80 37	10 20 93 70	3	A-C	No A are B No C are B	All A are C Some A are C No A are C Some A are not C	P P P P	10 27 43 3	60 63 77 77
3	C-A	No A are B All C are B	All C are A Some C are A No C are A Some C are not A	I I N N	3 10 83 40	17 13 87 57	3	C-A	No A are B No C are B	All C are A Some C are A No C are A Some C are not A	P P P P	17 20 40 27	57 57 77 67
3	A-C	No A are B Some C are B	All A are C Some A are C No A are C Some A are not C	P P P P	10 40 43 63	13 53 87 80	3	A-C	No A are B Some C are not B	All A are C Some A are C No A are C Some A are not C	P P P P	13 50 33 57	27 63 80 73
3	C-A	No A are B Some C are B	All C are A Some C are A No C are A Some C are not A	I P P N	7 37 40 60	3 57 63 80	3	C-A	No A are B Some C are not B	All C are A Some C are A No C are A Some C are not A	P P P P	3 43 30 50	13 60 77 70
3	A-C	Some A are not B All C are B	All A are C Some A are C No A are C Some A are not C	I P P N	3 77 7 90	17 87 40 90	3	A-C	Some A are not B No C are B	All A are C Some A are C No A are C Some A are not C	P P P P	3 33 27 53	27 80 50 83
3	C-A	Some A are not B All C are B	All C are A Some C are A No C are A Some C are not A	P P P P	7 50 27 80	17 80 43 90	3	C-A	Some A are not B No C are B	All C are A Some C are A No C are A Some C are not A	P P P P	7 47 30 67	30 63 70 67
3	A-C	Some A are not B Some C are B	All A are C Some A are C No A are C Some A are not C	P P P P	7 67 13 80	30 97 40 93	3	A-C	Some A are not B Some C are not B	All A are C Some A are C No A are C Some A are not C	P P P P	13 53 13 73	40 93 57 100
3	C-A	Some A are not B Some C are B	All C are A Some C are A No C are A Some C are not A	P P P P	3 60 13 63	27 93 30 96	3	C-A	Some A are not B Some C are not B	All C are A Some C are A No C are A Some C are not A	P P P P	3 57 13 60	43 100 33 87

Appendix B (continued)

Fig.	Conc. order	Premises	Conclusion	L	N	P	Fig.	Conc. order	Premises	Conclusion	L	N	P
4	A-C	All B are A All B are C	All A are C Some A are C No A are C Some A are not C	P N I P	70 27 10 10	87 70 27 43	4	A-C	All B are A No B are C	All A are C Some A are C No A are C Some A are not C	I P P N	7 20 63 30	20 17 93 70
4	C-A	All B are A All B are C	All C are A Some C are A No C are A Some C are not A	P N I P	60 47 10 17	87 70 27 37	4	C-A	All B are A No B are C	All C are A Some C are A No C are A Some C are not A	P P P P	7 20 73 37	13 27 90 67
4	A-C	All B are A Some B are C	All A are C Some A are C No A are C Some A are not C	P N I P	13 90 7 70	37 93 27 90	4	A-C	All B are A Some B are not C	All A are C Some A are C No A are C Some A are not C	I P P N	7 63 13 87	30 87 40 83
4	C-A	All B are A Some B are C	All C are A Some C are A No C are A Some C are not A	P N I P	10 83 10 60	43 90 27 37	4	C-A	All A are B Some C are not B	All C are A Some C are A No C are A Some C are not A	P P P P	0 53 17 83	20 80 27 87
4	A-C	Some B are A All B are C	All A are C Some A are C No A are C Some A are not C	P N I P	7 77 7 73	33 87 23 87	4	A-C	Some B are A No B are C	All A are C Some A are C No A are C Some A are not C	I P P N	7 30 30 70	23 60 80 83
4	C-A	Some B are A All B are C	All C are A Some C are A No C are A Some C are not A	P N I P	13 73 10 67	53 87 30 77	4	C-A	Some B are A No B are C	All C are A Some C are A No C are A Some C are not C	P P P P	7 30 47 53	20 60 77 83
4	A-C	Some B are A Some B are C	All A are C Some A are C No A are C Some A are not C	P P P P	7 57 17 60	47 90 60 90	4	A-C	Some B are A Some B are not C	All A are C Some A are C No A are C Some A are not C	P P P P	7 70 7 80	33 87 43 90
4	C-A	Some B are A Some B are C	All C are A Some C are A No C are A Some C are not A	P P P P	20 57 17 57	37 90 53 83	4	C-A	Some B are A Some B are not C	All C are A Some C are A No C are A Some C are not C	P P P P	0 57 13 67	33 87 47 93
4	A-C	No B are A All B are C	All A are C Some A are C No A are C Some A are not C	P P P P	0 17 63 53	17 20 83 60	4	A-C	No B are A No B are C	All A are C Some A are C No A are C Some A are not C	P P P P	20 17 47 23	60 50 80 77
4	C-A	No B are A All B are C	All C are A Some C are A No C are A Some C are not A	I P P N	10 27 63 33	7 17 87 60	4	C-A	No B are A No B are C	All C are A Some C are A No C are A Some C are not C	P P P P	23 30 43 27	53 57 80 73
4	A-C	No B are A Some B are C	All A are C Some A are C No A are C Some A are not C	P P P P	7 37 37 50	13 47 83 80	4	A-C	No B are A Some B are not C	All A are C Some A are C No A are C Some A are not C	P P P P	13 40 33 53	43 70 67 83
4	C-A	No B are A Some B are C	All C are A Some C are A No C are A Some C are not A	I P P N	3 40 27 57	10 57 77 77	4	C-A	No B are A Some B are not C	All C are A Some C are A No C are A Some C are not C	P P P P	13 37 43 50	23 57 63 77
4	A-C	Some B are not A All B are C	All A are C Some A are C No A are C Some A are not C	P P P P	13 67 17 87	23 83 50 87	4	A-C	Some B are not A No B are C	All A are C Some A are C No A are C Some A are not C	P P P P	13 43 33 50	37 60 67 70
4	C-A	Some B are not A All B are C	All C are A Some C are A No C are A Some C are not A	I P P N	7 60 13 87	40 80 23 90	4	C-A	Some B are not A No B are C	All C are A Some C are A No C are A Some C are not A	P P P P	13 37 40 47	57 57 70 83
4	A-C	Some B are not A Some B are C	All A are C Some A are C No A are C Some A are not C	P P P P	3 67 3 87	40 83 40 93	4	A-C	Some B are not A No B are C	All A are C Some A are C No A are C Some A are not C	P P P P	13 57 13 53	40 93 60 97
4	C-A	Some B are not A Some B are C	All C are A Some C are A No C are A Some C are not A	P P P P	3 53 13 67	27 80 60 83	4	C-A	Some B are not A Some B are not C	All C are A Some C are A No C are A Some C are not A	P P P P	3 53 17 60	43 93 53 90

Note. Fig. = figure; Conc. = conclusion; L = logic (under which N = necessary; I = impossible; P = possible).