The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pqja20

Reasoning from Suppositions

Ruth M.J. Byrne [a], Simon J. Handley [b] & Philip N. Johnson-Laird [c]

[a] Trinity College, University of Dublin, Dublin, Ireland

[b] University of Plymouth, Plymouth, U.K.

[c] Princeton University, Princeton, New Jersey, U.S.A.

Version of record first published: 29 May 2007.

To cite this article: Ruth M.J. Byrne , Simon J. Handley & Philip N. Johnson-Laird (1995): Reasoning from Suppositions, The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 48:4, 915-944

To link to this article: http://dx.doi.org/10.1080/14640749508401423

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 1995, *48A* (4), 915–944

Reasoning from Suppositions

Ruth M.J. Byrne
Trinity College, University of Dublin, Dublin, Ireland
Simon J. Handley
University of Plymouth, Plymouth, U.K.
Philip N. Johnson-Laird
Princeton University, Princeton, New Jersey, U.S.A.

Two experiments investigated inferences based on suppositions. In Experiment 1, the subjects decided whether suppositions about individuals' veracity were consistent with their assertions—for example, whether the supposition "*Ann is telling the truth and Beth is telling a lie*", is consistent with the premises: "*Ann asserts: I am telling the truth and Beth is telling the truth. Beth asserts: Ann is telling the truth*". It showed that these inferences are more difficult than ones based on factual premises: "*Ann asserts: I live in Dublin and Beth lives in Dublin*". There was no difference between problems about *truthtellers* and *liars*, who always told the truth or always lied, and *normals*, who sometimes told the truth and sometimes lied. In Experiment 2, the subjects made inferences about factual matters set in three contexts: a truth-inducing context in which friends confided their personality characteristics, a lie-inducing context in which business rivals advertised their products, and a neutral context in which computers printed their program characteristics. Given the supposition that the individuals were lying, it was more difficult to make inferences in a truth-inducing context than in the other two contexts. We discuss the implications of our results for everyday reasoning from suppositions, and for theories of reasoning based on models or inference rules.

Reasoning in daily life often depends on supposing that some proposition holds and then drawing out the consequences of this supposition. Suppose—you say to yourself—that inflation increases as the economy picks up, what then? And you use your general knowledge to infer various consequences—for example, that your savings will lose value. In this sort of case, your conclusion should embody your assumption. You should

Requests for reprints should be sent to R.M.J. Byrne, Department of Psychology, Trinity College, University of Dublin, Dublin 2, Ireland. Email: rmbyrne@vax1.tcd.ie; Fax: +353–1–612006.

This research was funded by an ESRC grant, Reference R000 23 2491, awarded to Ruth Byrne, and by support from the James S. McDonnell Foundation to P.N. Johnson-Laird. We thank Mark Keane for suggesting the idea of the second experiment, Alan Milne for statistical help, and Henry Markovits and Mike Oaksford for helpful comments on an earlier draft. We are grateful to the School of Psychology, University of Wales College of Cardiff, for providing the facilities to conduct the research.

conclude: if inflation increases, then my savings will lose value. In other cases, your conclusion may feed back on your supposition and show it to be false. Consider, for example, the following sort of inference (Evans, 1972):

You must either go to the pictures tonight or go for a walk tomorrow (but not both).
If you go to the pictures tonight, you must go for a walk tomorrow.
What follows?

You can argue by making a supposition: Suppose that I go to the pictures tonight, then it follows from the conditional that I must go for a walk tomorrow. But the first premise asserts that going to the pictures tonight rules out going for a walk tomorrow. Hence, my supposition is false: I cannot go to the pictures tonight (on pain of contradiction). This inference, in turn, leads from the first premise to the further conclusion that I must go for a walk tomorrow.

The general strategy of *suppositional* reasoning, as we will refer to the process, calls for at least two stages: (1) making a supposition—that is, making an assumption on a hypothetical basis; and (2) drawing some inferential consequences from the assumption. Psychologists have amassed considerable information on the second stage—that is, how reasoning proceeds from premises taken to be true (for a review, see Evans, Newstead, & Byrne, 1993). They have amassed less information about the first stage. They have enquired into it and, in particular, into the circumstances in which individuals are likely to make a supposition. One conjecture is that they will do so when there is no definite information—no categorical premises—on which to start the inferential process, as in the example above about the pictures and the walk. Psychologists have also enquired into what particular proposition is likely to be the basis for a supposition, and there is some evidence that, if possible, it will be a causal antecedent. Thus, consider the premises:

If Bill doesn't stay sober, then he doesn't keep to his diet.
Either he keeps to his diet or else he gets depressed.

In this case, reasoners are likely to make the supposition that Bill does not stay sober, and they infer that he does not keep to his diet, and then that he gets depressed. In short, if Bill does not stay sober, then he gets depressed. However, given the premises in the following guise:

Bill doesn't stay sober only if he doesn't keep to his diet.
Either he keeps to his diet or else he gets depressed.

the same conclusion follows validly, but the task is more difficult (Johnson-Laird & Shapiro, see Wason & Johnson-Laird, 1972, p. 74). The first premise now makes salient the causal possibility that Bill does not stay on his diet, but this merely leads to a conclusion equivalent to the second premise. It is necessary to make the supposition that Bill does not stay sober, from which it follows that he did not keep to his diet and hence gets depressed.

What mental mechanisms underlie suppositional inferences? One possibility is that they rely on formal rules of inference akin to those of a logical calculus. This idea is highly plausible, as logical systems based on the principles of *natural deduction* (see Gentzen, 1935; Prawitz, 1965) include a rule of conditional inference: If a proposition, *p*, is the sole supposition in a derivation yielding the conclusion, *q*, then one can draw the conclusion, *if p then q*. As Quine (1974, p. 207) has remarked, this rule is the crux of natural deduction, and it is indeed a feature of those psychological theories of deductive reasoning based on formal rules of inference (see, e.g., Braine & O'Brien, 1991; Rips, 1994). These theories also recognize that the chain of inferences leading from a supposition can demonstrate that the supposition is false. The theories, accordingly, include a rule of *reductio ad absurdum*: If a proposition, *p*, is the sole supposition in a derivation that yields a self-contradiction (in which a proposition, *q*, and its negation, *not-q*, both hold), then one can conclude: *not p*.

An alternative theoretical possibility is that individuals reason by constructing mental models. They understand what the premises mean, and they use this understanding to construct a model of the situation that they describe (Johnson-Laird & Byrne, 1991). A conclusion is valid provided that it is true in any such model. This theory, like those based on formal rules, draws a distinction between inferential strategy and the mechanism for making inferences. Thus, reasoners can make suppositional inferences: They make a hypothesis and then use model-based procedures for inferring consequences from it and the premises.

Our aim in the present paper is to advance the understanding of how reasoners make inferences on the basis of suppositions, and of how the content and the context of premises influence the inferential process. We will compare theories based on formal rules with the mental model theory. The two sorts of theories offer competing explanations of suppositional inferences, and their predictions diverge for an important class of inferences that we describe in the next section.

Meta-inferences, Rules, and Models

Certain inferences concern, not direct matters of fact, but the consistency of assertions made by different individuals. As an illustration, consider this case, where two friends confide in each other their assessment of some of their personality characteristics:

Anne asserts: I am efficient and Beth is not efficient.
Beth asserts: Anne is not efficient.

Both Anne and Beth may believe that they are each telling the truth, and their beliefs may even be supported by different criteria for efficiency—for example, Anne's assertion may be supported by her belief that she is efficient most of the time, whereas Beth's belief may be supported by her belief that Anne is inefficient on occasions. Putting aside such equivocations, their two assertions cannot both be true (using either objective criteria for establishing efficiency, or personal criteria). Suppose, for example, that Anne's assertion is true; then it follows that Anne is efficient. Suppose, further, that Beth's assertion is true; then it follows that Anne is not efficient. Hence, both assertions cannot be true. You

can readily infer that either Anne's assertion is true and Beth's assertion is not, or vice versa. But what about the possibility that neither Anne's nor Beth's assertion is true? You might find it more difficult to work out that this possibility is also consistent with the premises. Anne could be right about herself, but her conjunctive assertion could nevertheless be false because Beth is efficient; and Beth's assertion could be false because Anne is indeed efficient. Hence, you can infer that Anne and Beth cannot both be telling the truth, but not which of them is telling the truth, or whether either of them is. In Table 1, we illustrate the set of relevant paths to pursue in order to evaluate each possibility (see Byrne & Handley, 1993; Jeffrey, 1981). These diagrams represent the set of possibilities that must be considered, but the actual mental processes on which reasoners rely may be very different.

A special subclass of inferences about consistency concern individuals' claims about the truth or falsity of their own and others' assertions; for example:

Anne asserts: I am efficient.
Beth asserts: That's a lie.

These meta-inferences—inferences about the truth and falsity of individuals' assertions—are central to interactions among people in social, political, business, legal, and computing environments (e.g. Anno & Nozaki, 1984; Dewdney, 1989; Fujimara, 1884; Goffman, 1959). They readily yield deep paradoxes:

Anne asserts: Beth's next assertion is true.
Beth asserts: Anne's last assertion is false.

and are accordingly central to the foundations of modern logic (e.g. Austin, 1970; Barwise & Etchemendy, 1987; Kripke, 1975; Tarski, 1944). Such dialogues can also yield logical puzzles in the guise of *truthteller–liar* puzzles, which were exploited in entertaining and instructive ways by Smullyan (e.g. 1978). These puzzles typically concern two sorts of individuals: those who always tell the truth (*truthtellers*, a.k.a. *knights*) and those who always lie (*liars*, a.k.a. *knaves*), and the task is to infer the status of the various individuals on the basis of what they have to assert about one another. An example of such a puzzle is presented in Table 2, along with its logical analysis and solution. The reader should appreciate, however, that the same sorts of inference can arise in circumstances where contingent claims are made about the truth or falsity of assertions. The problem in Table 2, for example, is logically equivalent to the one raised by the following dialogue between two individuals who may or may not be telling the truth:

Anne asserts: I am asserting the truth in this assertion, or else Beth is lying in what she asserts, but not both.
Beth asserts: Anne is lying.

Psychologists have recently begun to investigate *truthteller–liar* puzzles (Byrne & Handley, 1992, 1993; Byrne, Handley, & Johnson-Laird, 1992; Byrne, Johnson-Laird, & Handley, 1993; Evans, 1990; Johnson-Laird & Byrne, 1990, 1991; Rips, 1989, 1990).

TABLE 1
Tree Diagrams to Illustrate the Relevant Paths to Evaluate Alternative Possibilities

The premises:
Anne says: Anne is efficient and Beth is not efficient.
Beth says: Anne is not efficient.

Supposition 1. *Anne is telling the truth, and Beth is telling the truth*:

$A_{truthtelling}$ $B_{truthtelling}$
| |
$A_{efficient}$ $A_{not\text{-}efficient}$ *Inconsistent*
$B_{not\text{-}efficient}$

The first row contains "$A_{truthtelling}$", which corresponds to the supposition that Anne is telling the truth and her assertion is true. If so, a single path follows—represented vertically in the diagram—A is efficient and B is not efficient. The first row also contains "$B_{truthtelling}$", which corresponds to the supposition that Beth is telling the truth and her assertion is true. In this case, the path that follows is: A is not efficient. This possibility contradicts the one from the supposition that A is telling the truth. Because the two paths lead to a contradiction, Supposition 1 is false.

Supposition 2: *Anne is telling the truth and Beth is lying:*

$A_{truthtelling}$ B_{lying}
| |
$A_{efficient}$ $A_{efficient}$ *Consistent*
$B_{not\text{-}efficient}$

Because the two paths are consistent, Supposition 2 may be true.

Supposition 3. *Anne is lying and Beth is telling the truth:*

A_{lying} $B_{truthtelling}$
| |
$A_{efficient}$ $A_{not\text{-}efficient}$ $A_{not\text{-}efficient}$ $A_{not\text{-}efficient}$ *Consistent*
$B_{efficient}$ $B_{not\text{-}efficient}$ $B_{efficient}$

"A_{lying}" corresponds to the supposition that A is lying and so her assertion is false. This supposition leads to three alternative paths: If A's assertion is false, both A and B are efficient (the first possibility), neither of them is efficient (the second possibility), or A is not efficient and B is (the third possibility). The first of these possibilities is inconsistent with the supposition that B's assertion is true, but the remaining two are consistent. Hence, Supposition 3 may be true.

Supposition 4. *A is lying and B is lying*

A_{lying} B_{lying}
| |
$A_{efficient}$ $A_{not\text{-}efficient}$ $A_{not\text{-}efficient}$ $A_{efficient}$ *Consistent*
$B_{efficient}$ $B_{not\text{-}efficient}$ $B_{efficient}$

The first of the possibilities given that A's assertion is false is consistent with B's assertion being false. Hence, Supposition 4 may be true.

They have been used to study deductive reasoning, much as puzzles such as the Tower of Hanoi have been used to study problem solving—not because they are likely to be encountered in everyday life, but because they are informative about underlying mental processes (see, e.g. Keane, Ledgeway, & Duff, 1994; Kotovsky & Simon, 1989; Newell,

TABLE 2
An Example of a Truthteller–Liar Puzzle and the Paths to Solve It

The premises:

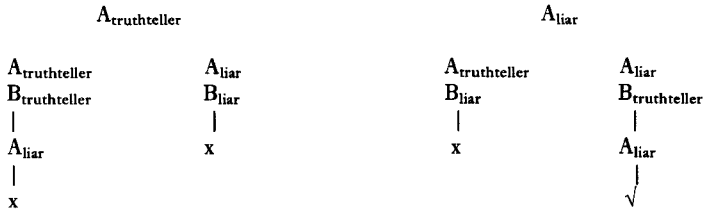
Imagine an island inhabited by truthtellers and liars. Truthtellers always tell the truth, and liars always lie. You overhear the following conversation between two of the inhabitants:

A asserts: I am a truthteller or else B is a liar, but not both.

B asserts: A is a liar.

Is A a truthteller, or a liar, and is B a truthteller or a liar, or is there insufficient information to know?

The complete set of relevant paths to pursue to solve this puzzle:



The first insertion in the first row corresponds to the supposition that A is a truthteller and her assertion is true. If so, two alternative paths follow—A and B are both truthtellers, or they are both liars. The first of these paths leads to the possibility that A is a liar because B is telling the truth and she says A is a liar. This possibility contradicts the initial supposition that A is telling the truth, as represented with an "x". The second path, from the possibility that A and B are both liars, immediately leads to a contradiction, because the possibility that A is a liar conflicts with the initial supposition that she is telling the truth. Hence both paths from the supposition that A is telling the truth lead to a contradiction, and so the supposition is false.

The second set of paths follows from the supposition that A is a liar and her assertion is false. If so, then A is a truthteller and B is a liar, or vice versa. The first of these possibilities leads immediately to a contradiction. The second path leads to the possibility that A is a liar, because B says she is, and thus to a consistent assignment: A is a liar and B is a truthteller, represented with a "√". Because the supposition is the only one that leads to a consistent assignment, it must be true.

1990; Newell & Simon, 1972). A theory of *truthteller–liar* puzzles and of meta-inference in general needs to account both for the inferential strategies that reasoners use, such as making a supposition, and for the mechanisms by which they make individual inferential steps, such as inferences based on sentential connectives (e.g. "and", "or", "if") and on truth and falsity. In fact, there are two competing theories of meta-inference, based on formal rules of inference and the other based on mental models.

The rule theory (Rips, 1989, 1990) postulates that reasoners rely on a single suppositional strategy to solve meta-inferential problems. They make the supposition that the first individual in a puzzle is telling the truth and they draw as many inferences as possible using a mental repertoire of formal rules; they can also make the supposition that the first individual is telling a lie and pursue the inferential consequences of this assumption. This strategy is similar to the one we have used in Table 2. The theory postulates that individual inferential steps are made on the basis of formal rules of inference. Thus, inferences based on sentential connectives, such as:

A is a truthteller or B is a truthteller.
A is not a truthteller.
Therefore, B is a truthteller.

are derived using the formal rule:

p OR *q* NOT-*p* Therefore, *q*

(Rips, 1989, p. 94; see also Rips, 1983). The theory also postulates certain content-specific rules for making inferences about truth (or truthtellers) and falsity (or liars). Thus, in order to derive the following sort of inference:

A asserts: B is a liar.
A is a truthteller.
Therefore, B is a liar.

the theory adopts the following sort of rule:

says (x, p)
truthteller (x) Therefore, p.

(Rips, 1989, p. 90). The second clause of the rule may refer instead to *knight*(x) or to tells-the-truth(x), or to any other device that represents x as a source of truth. Likewise, in order to derive the following sort of inference:

A asserts: B is a liar.
A is a liar.
Therefore, B is not a liar.

the theory adopts the following rule:

says (x, p)
liar(x) Therefore, not p.

The second clause of the rule may refer instead to *knave*(x) or to any other device that represents x as a source of a false assertion. The theory appears to be supported by the observation that subjects make more errors and take longer to solve problems that require more inferential steps in their derivations (Rips, 1989).

In contrast, the model theory proposes that logically naive individuals do not possess a single uniform strategy for making meta-inferences. Indeed, they have no robust *a priori* strategies for dealing with these puzzles, but have to learn how to deal with them. Hence, subjects are likely to discover specific strategies geared to the particular nature of the puzzles presented to them. Consider, for example, a problem of the form:

A asserts: C is a truthteller.
B asserts: C is a truthteller.
C asserts: A is a truthteller and B is a liar.

Reasoners are likely to notice that both A and B make the same assertion, whereas C characterizes them as having a distinct status (one a *truthteller* and the other a *liar*). But A and B must have the same status, because they make the same assertion, and so C cannot be telling the truth. Hence, as C is a *liar*, then so too are A and B, because they both assert that C is a *truthteller*. This strategy, and other simple strategies that we have modelled computationally (see Johnson-Laird & Byrne, 1990), minimize the number of possibilities that reasoners need to keep in mind in order to reach a solution.

In contrast to the uniform strategy proposed by Rips, the present account is corroborated by the observation that reasoners make more correct inferences to problems that can be solved by these simple strategies than to problems that cannot be solved by them (Byrne and Handley, forthcoming). In order to make inferential steps based on sentential connectives, the model theory postulates that reasoners construct appropriate sets of models for each sort of connective (Johnson-Laird, Byrne, & Schaeken, 1992, p. 422). For example, the assertion:

A is a truthteller or B is a truthteller, or both

calls for models of three possibilities, which are represented in the following diagram:

```
A
      B
A     B
```

where each line denotes a separate model, and "A" represents A as a *truthteller* and "B" represents B as a *truthteller*. Such models are partially implicit in that they do not represent explicitly that in the first model B is not a *truthteller*, or that in the second model A is not a *truthteller*. Implicit models reduce the load on the processing capacity of working memory, and they also have empirical consequences that have been corroborated experimentally (see Johnson-Laird & Barres, 1994). If necessary, however, reasoners can flesh out implicit models in a wholly explicit way:

```
 A     −B
−A      B
 A      B
```

where "−" represents negation, and so −A represents A as not a *truthteller* (i.e. A is a *liar*). Finally, in order to cope with inferences about truth (and *truthtellers*) and falsity (and *liars*), the theory postulates that reasoners can annotate their models to indicate that they correspond to assertions made by individuals (Johnson-Laird & Byrne, 1990, p. 73). For example, the following premise:

A asserts: I am a truthteller or B is a truthteller, or both.

is represented by the following models:

A: A
B
A B

where the annotation "A:" shows that the models correspond to the content of A's assertion.

The two sorts of theory—one syntactic (formal rules) and the other semantic (mental models)—may not exhaust the theoretical possibilities, but they offer the only current accounts of meta-inference. The goal of the present study is, accordingly, to compare their accounts of the inferences that naive reasoners make about the truth of assertions.

EXPERIMENT 1

The Effects of the Assertor's Reliability and the Assertion's Content

In previous psychological studies of meta-inference, the task has been to identify the status (as *truthteller* or *liar*) of each individual in a problem; for example, "Is Anne telling the truth or lying, is Beth telling the truth or lying, or is there insufficient information to know?" Such questions may call for the pursuit of many alternative inferential paths, as Table 2 shows. One consequence has been that subjects perform poorly and tend to develop idiosyncratic strategies. We therefore created a new and simpler inferential task by breaking the original complex one into smaller separate suppositional components. We presented the subjects with each of the possible suppositions separately; for example:

Anne is telling the truth and Beth is telling the truth.
Anne is telling the truth and Beth is lying.
Anne is lying and Beth is telling the truth.
Anne is lying and Beth is lying.

and they had to judge whether each supposition was consistent or inconsistent with the premises, such as:

Anne asserts: I am telling the truth and Beth is telling the truth.
Beth asserts: Anne is lying.

Each possibility is a supposition, and reasoners find it easier to make suppositional inferences when they are prompted to make the initial supposition (Byrne & Handley, submitted).

The aim of Experiment 1 was to examine the influence of information about the reliability of the assertor. Previous studies used problems based on individuals who either

always tell the truth (*truthtellers* or *knights*) or always lie (*liars* or *knaves*). Such individuals are unusual in everyday life, where most people sometimes tell the truth and sometimes lie. Our experiment compared inferences from the two sorts of premises: problems based on *truthtellers* and *liars*, and problems based on *normals*, who sometimes tell the truth and sometimes lie. There have been no studies of the cognitive processes used to make inferences about normal assertors, and it is possible that the inferences about *truthtellers* and *liars* are not representative of everyday reasoning about truth. Indeed, the concepts of *truthteller* and *liar* may be so artificial and unfamiliar that they fail to engage the subjects' everyday reasoning abilities. We took care to ensure that the problems based on normal assertors were formally isomorphic to those based on *truthtellers* and *liars*, i.e. the two sorts of problems had the same structure, the same number of solution paths, and the same number of consistent solutions. The isomorphism depended on using identical assertions in both conditions:

Anne says: I am telling the truth and Beth is lying.

which were presented in a context that made clear either that the individuals were *truthtellers* or *liars*, or else that they were *normals*. In either case, the chain of inferences required determine whether a supposition is consistent or inconsistent with the premises is identical. On the one hand, if the difficulty of *truthtellers–liars* puzzles arises from their artificiality or unfamiliarity, then puzzles based on *normal* individuals should be easier. On the other hand, if the difficulty of the puzzle arises from the complexities of their logic, then there will be no reliable difference between the two sorts of problem, because they have the same underlying logic.

A second aim of the experiment was to compare the model theory with the rule theory by examining the influence of the kinds of information that the individuals in the problems assert. Previous studies of meta-inference have called for inferences from assertions about *truthtellers* and *liars* to conclusions of the same sort. In everyday life, however, conclusions about the veracity of individuals tend to be based on premises about matters of fact. Hence, the experiment compared inferences from premises about truth status (e.g. *Anne is telling the truth*) with inferences from factual premises (e.g. *Anne lives in Paris*). In both cases, the conclusions concerned the truth-telling status of the individuals. Rule theories and the model theory make different predictions about the two sorts of problems. Consider, for instance, how the rule theory deals with the following problem:

A asserts: I am telling the truth, and B is telling the truth.
B asserts: A is telling the truth.
Supposition to be evaluated: *A is telling the truth and B is telling a lie.*

Given the suppositional strategy and the sorts of rules of inference postulated by Rips (1989), the derivation proceeds as follows (where we include the content of the problem for clarity):

i.	A is telling the truth	[conjunction elimination, from supposition]
ii.	B is telling a lie	[conjunction elimination, from supposition]
iii.	A is not telling the truth	[rule for *liar*, from ii and second premise]

iv. A is telling the truth and A is not telling the truth
[conjunction introduction, i and iii]
v. not (A is telling the truth, and B is telling a lie)
[reductio ad absurdum, contradiction in iv, and supposition]

The corresponding factual problem is:

A asserts: A lives in Dublin, and B lives in Dublin.
B asserts: A lives in Dublin.
Supposition to be evaluated: *A is telling the truth and B is telling a lie.*

The derivation for this problem is similar to the previous one, but two steps longer:

i. A is telling the truth [conjunction elimination, from supposition]
ii. B is telling a lie [conjunction elimination, from supposition]
iii. A does not live in Dublin
[rule for *liar*, from ii and second premise]
iv. A lives in Dublin, and B lives in Dublin
[rule for *truthteller*, from i and first premise]
v. A lives in Dublin [conjunction elimination, from iv]
vi. A lives in Dublin and A does not live in Dublin
[conjunction introduction, iii, v]
vii. not (A is telling the truth and B is telling the truth)
[reductio ad absurdum, contradiction in vi, and supposition]

The two extra steps (iv and v) are needed to recover the content of A's assertion, which (unlike the truth-status inference) is not given in the supposition. These steps are necessary in such inferences, and so the rule theory predicts that these sorts of factual problems should be more difficult than the truth-status inferences. We have devised a computer program that implements the model theory's account of the suppositional strategy. Given the truth-status problem:

A asserts: I am telling the truth and B is telling the truth.
B asserts: A is telling the truth.
Supposition to be evaluated is: *A is telling the truth and B is telling a lie.*

the program returns the following output:

```
SUPPOSE FROM SUPPOSITION S1 THAT A AND NOT B.

SUPPOSE     hyp A:        A     B
            neg-hyp B:    -B    -A
INCONSISTENT
```

where the first line represents the supposition (presented by the experimenter), the second line represents the first clause of the supposition (A is telling the truth) and its immediate consequence, and the third line represents the second clause of the supposition (B is lying, i.e. not telling the truth) and its immediate consequence. The two assumptions yield models that are inconsistent, and so the supposition is inconsistent with the premises. Likewise, the factual problem:

A asserts: A lives in Dublin and B lives in Dublin.
B asserts: A lives in Dublin.
Supposition to be evaluated: *A is telling the truth, and B is telling a lie.*

can be solved as follows:

hyp A:	A	A_{dublin}	B_{dublin}
neg-hyp B:	$-B$	$-A_{dublin}$	

which also yield inconsistent models. The theory allows that subjects may have recourse to other strategies, but there is a danger in all the truth-status tasks of confusing the annotation on a model with its content—that is, people may confuse the supposition that A is telling the truth with A's assertion that A is telling the truth. This difficulty does not arise in the case of factual assertions, because their content (e.g. A lives in Dublin) is quite remote from that of the supposition that A is telling the truth. With assertions about truth (see Table 3) we show disjunctive consequences that require three alternatives to be followed up. In fact, the program immediately eliminates those alternatives that conflict with the status of the assertor—this process is required only with assertions about truth status, not with assertions about location. Hence the elimination of one of the three disjunctive alternatives is unique to the truth status problems. In summary, the model theory predicts that the factual problems should be easier than the truth-status problems. Hence, the two theories make opposing predictions about the relative difficulty of the two sorts of problems.

Method

Materials and Design. We systematically manipulated the two variables to produce four sorts of problems with the same suppositions:

1. problems concerning *truthtellers* and *liars*, who referred to the truth-status of their own and the other's assertions—e.g. *I am telling the truth*;
2. problems concerning *normal* individuals, who referred to the truth-status of their own and the other's assertions—e.g. *I am telling the truth*;
3. problems concerning *truthtellers* and *liars*, who referred to the places where they lived—e.g. *I live in Dublin*;
4. problems concerning *normal* individuals, who referred to the place where they lived—e.g. *I live in Dublin.*

Table 3 presents an example of each of the four sorts of problems, the four suppositions that the subjects had to evaluate, and a logical analysis of the task. Each problem contained two premises, and each premise was an assertion made by one of two speakers, who were identified by single-syllable names of the same gender, none of which began with the same initial letter. The first premise in each problem contained A's assertion, which consisted of two clauses about A and B. It was one of eight sorts, depending on whether the connective was a conjunction (A and B) or an inclusive disjunction (A or B, or both), and on whether each constituent proposition was affirmative or negative (A, not-A, B, not-B). The second premise contained B's assertion, which consisted of a single clause about A, which was either affirmative or negative (A, not-A). The alternative possibilities gave rise to 16 (8×2) distinct sorts of problems. Each of these problems was presented with four suppositions for the subjects to evaluate as consistent or inconsistent with the premises:

A is telling the truth and B is telling the truth.
A is telling the truth and B is lying.
A is lying and B is telling the truth.
A is lying and B is lying.

We constructed two sets of 32 problems, one set based on *truthtellers* and *liars*, that is, individuals who always tell the truth or always lie, and the other set based on *normals*, that is, individuals who sometimes tell the truth and sometimes lie. Within each set of problems, 16 problems contained assertions based on truth-status, e.g. "*I am telling the truth, and B is telling the truth*", and 16 contained assertions based on the factual matter of location, e.g. "*I live in Dublin, and B lives in Dublin*". (The full set of problems is presented in Appendix A along with the numbers of correct responses.) We gave one group of subjects the *truthtellers–liars* set, and another group of subjects the *normals* set. Each subject completed the 16 problems based on truth-status, and the 16 problems based on location, presented in separate blocks. Half the subjects in each group received the truth-status problems first followed by the location problems, and the other half received the problems in the opposite order. Each subject received the problems in a different random order within each block.

Procedure. The subjects were tested in several medium-sized groups, and they were randomly assigned to one of the two groups. They were told that the experiment was designed to examine ordinary reasoning and was not a test of intelligence. They were also told that most people found the problems difficult, but that they should take their time to think through each of them fully. They were given a practice problem, which was based on *normals* for the normal group, and *truthtellers–liars* for the truthtellers–liars group. The experimenter explained that their task was to evaluate four possible states of affairs (the suppositions) as consistent, or inconsistent, with the premises. Each problem was presented on a separate sheet of paper, with the four suppositions listed below the premises. The subjects made their responses by placing a tick either under the heading "consistent" or else under the heading "inconsistent" beside each supposition. They were asked to work at their own pace, to answer the problems in the order that they were given, and not to return to a problem once they had completed it.

Subjects. Thirty-two subjects (19 women and 13 men), undergraduate students and members of the subject panel from the University of Wales College of Cardiff, were paid £3 per hour for their participation in the experiment, which lasted approximately 40 minutes. The subjects, whose ages ranged from 18 to 34, had no formal training in logic and had not previously participated in an experiment on reasoning. They were randomly assigned to one of two groups, either the truthtellers–liars or the normals group ($n = 16$ in each).

TABLE 3
The Four Sorts of Problems Used in Experiment 1 and Diagrams to Illustrate the Conclusions That Follow from Them

The premises of two sorts of problems:

1. *Imagine an island inhabited by truthtellers, that is, people who always tell the truth, and liars, that is, people who always lie.*
   Anne says: I am telling the truth and Beth is lying
   Beth says: Anne is lying

2. *Imagine an island inhabited by normal people, that is, people who sometimes tell the truth and who sometimes lie.*
   Anne says: I am telling the truth and Beth is lying
   Beth says: Anne is lying

The possibilities can be judged in the following way for both Problems 1 and 2:

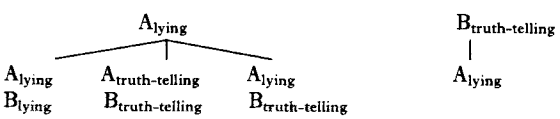
Supposition 1 *A is telling the truth and B is telling the truth:*

Inconsistent

Supposition 2 *A is telling the truth and B is lying:*

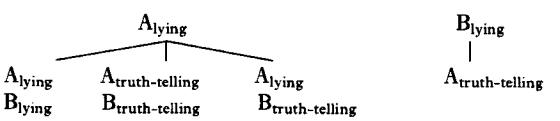
Consistent

Supposition 3 *A is lying and B is telling the truth:*



Consistent

Supposition 4 *A is lying and B is lying:*



Inconsistent

Results

Table 4 summarizes the results, which are presented in detail in Appendix A. An analysis of variance (with the between-subject factor of the assertor's status and two within-subject factors—the assertion's content and the supposition type) showed that there was no reliable difference between the truthtellers–liars group (65% correct inferences) and the normals group (67% correct inferences), $F(1, 30) = 0.31$, $p = 0.58$. However, there were more correct inferences for the problems based on location (74%) than for the

TABLE 3
(Continued)

The premises of two sorts of problems:

3. *Imagine an island inhabited by truthtellers, that is, people who always tell the truth, and liars, that is, people who always lie.*
   Anne says: I live in Dublin and Beth lives in Paris.
   Beth says: Anne lives in Paris.

4. *Imagine an island inhabited by normals, that is, people who sometimes tell the truth and sometimes lie.*
   Anne says: I live in Dublin and Beth lives in Paris.
   Beth says: Anne lives in Paris.

The possibilities can be judged in the following way for both Problems 3 and 4:

Supposition 1' *A is telling the truth and B is telling the truth:*

$A_{truth\text{-}telling}$ — $B_{truth\text{-}telling}$
|
A_{dublin} — A_{paris} *Inconsistent*
B_{paris}

Supposition 2' *A is telling the truth and B is lying:*

$A_{truth\text{-}telling}$ — B_{lying}
|
A_{dublin} — $A_{not\text{-}paris}$ *Consistent*
B_{paris}

Supposition 3' *A is lying and B is telling the truth:*

A_{lying} — $B_{truth\text{-}telling}$
|
A_{paris} *Consistent*

$A_{not\text{-}dublin}$ A_{dublin} $A_{not\text{-}dublin}$
B_{paris} $B_{not\text{-}paris}$ $B_{not\text{-}paris}$

Supposition 4' *A is lying and B is lying:*

A_{lying} — B_{lying}
|
$A_{not\text{-}paris}$ *Consistent*

$A_{not\text{-}dublin}$ A_{dublin} $A_{not\text{-}dublin}$
B_{paris} $B_{not\text{-}paris}$ $B_{not\text{-}paris}$

problems based on truth status (59%), $F(1, 30) = 17.42$, $p = 0.000$. The two variables exhibited a marginally reliable interaction, $F(1, 30) = 3.63$, $p = 0.067$. The type of supposition also yielded a reliable effect, $F(3, 90) = 3.47$, $p < 0.001$, i.e. of the task of evaluating whether A and B were telling the truth (70% correct), A was telling the truth and B was lying (67% correct), A was lying and B was telling the truth (66% correct), or A and B were both lying (63% correct). This variable interacted with the assertor's status, $F(3, 90) = 3.25$, $p < 0.02$, and with the assertion's content, $F(3, 90) = 2.95$, $p < 0.03$.

TABLE 4
The Percentage of Correct Conclusions in the Conditions in Experiment 1

		Truth-status	*Location*	*Total*
truthteller–liars	A, B	63	73	68
	A, not-B	64	67	66
	not-A, B	64	72	68
	not-A, not-B	53	63	58
	total	*61*	*69*	*65*
normals	A, B	64	77	71
	A, not-B	56	78	67
	not-A, B	55	70	63
	not-A, not-B	47	86	67
	total	*56*	*78*	*67*
	overall total	*59*	*74*	

Note: A, B — A is telling the truth and B is telling the truth
A, not-B — A is telling the truth and B is lying
not-A, B — A is lying and B is telling the truth
not-A, not-B — A is lying and B is lying

The subjects' task was to judge each of these options as consistent or inconsistent with the premises.

There was a reliable three-way interaction, $F(3, 90) = 2.73$, $p < 0.04$. Simple-effects analyses showed that subjects in the normals group made more correct inferences for the problems based on location than for the problems based on truth status, and the difference was reliable for every supposition type: A and B telling the truth (77% vs. 64%), $F(1, 120) = 3.66$, $p < 0.05$; A telling the truth and B lying (78% vs. 56%), $F(1, 120) = 10.61$, $p < 0.001$; A lying and B telling the truth (70% vs. 55%), $F(1, 120) = 4.88$, $p < 0.02$); and A and B both lying (86% vs. 47%), $F(1, 120) = 32.91$, $p < 0.000$. In contrast, for the truthtellers–liars group, the difference between the location and the truth problems did not reach significance for any supposition type: A and B telling the truth (73% vs. 63%), $F(1, 120) = 2.46$, $p = 0.11$); A telling the truth and B lying (67% vs. 64%), $F(1, 120) = 0.27$, $p = 0.6$; A lying and B telling the truth (72% vs. 64%), $F(1, 120) = 1.64$, $p < 0.2$; and A and B both lying (63% vs. 53%), $F(1, 120) = 2.29$, $p < 0.13$.

Overall, as we have seen, there was no difference between the normals and the truthtellers–liars groups. Indeed, for truth status problems, there was no reliable difference between them for any of the four sorts of supposition (see Table 4); for location problems, there was no reliable difference between them, except for one sort of supposition: A and B both lying (86% vs. 63%), $F(1, 240) = 12.99$, $p < 0.000$. Finally, as a glance at the data in Appendix A shows, the problems based on conjunctions were easier overall than the problems based on disjunctions, both for the truthtellers–liars group (69% vs. 61%), Wilcoxon's $t = 12$, $n = 16$, $p < 0.01$, and for the normals group (72% vs. 61%) Wilcoxon's $t = 1$, $n = 14$, $p < 0.001$.

Discussion

Even though *truthtellers–liars* problems are unlikely to be encountered in daily life, they were not reliably more difficult than problems about individuals who sometimes tell the truth and sometimes lie (the *normals* problems). Hence, the study of suppositional inferences about *truthtellers* and *liars* may be more generalizable to everyday assessments of truth than appears at first glance. However, as the model theory predicts, reasoners did find it more difficult to make inferences about truth status (e.g. "*Beth is telling the truth*"), than to make inferences about factual assertions (e.g. "*Beth lives in Dublin*"). Inferences about truth status, according to the model theory, call for reasoners to keep track of both an assertor's status (e.g. suppose A is telling the truth) and the content of his or her assertion (e.g. I am telling the truth), and to eliminate any contradictions that may arise between them. Because of the similarity in content, subjects are likely to confuse the two. The distinction between an assertor's truth status and the content of an assertion is much greater in the case of the factual problems, and so these problems are easier. This effect was greater for the normals group than for the truthtellers–liars group. It may be more difficult for reasoners to keep track of the compatibility of normal assertors and their assertions, because the current status of an individual who sometimes tells the truth and sometimes lies is more uncertain than the current status of someone who either always tells the truth or else always lies.

In short, the status of the individuals as *truthtellers–liars* or *normals* did not appear to affect performance overall, but the content of their assertions did affect performance: the task was easier with factual materials than with truth-status materials. Hence, there is an effect of the content of suppositional inferences.

EXPERIMENT 2

The Effects of Context

Ordinarily, when you make an assessment of truth, you do not do so in a vacuum. Your knowledge about the speakers, their goals, and the situation in which they make their assertions are all likely to influence you. The model theory postulates that such effects occur because reasoners' models embody their background knowledge and any other available relevant information. Experimental evidence corroborates this claim. Indeed, context influences inferences in a wide range of situations (for reviews, see Eysenck & Keane, 1990, Chapters 11 and 12; Evans, Newstead, & Byrne, 1993, Chapters 2, 3 and 4). Thus, for example, background knowledge affects whether reasoners assume that several conditions hold conjointly or as disjunctive alternatives to each other (Byrne, 1989a, 1989b; Byrne & Johnson-Laird, 1992). It can also affect the likelihood that reasoners will flesh out their initial models to be fully explicit and the ease with which they envisage a counterexample in Wason's selection task (Johnson-Laird & Byrne, 1991, Chapter 4).

Our aim in Experiment 2 was to examine the influence of context on meta-inferences based on suppositions. Our hypothesis was that some contexts, such as a discussion

among friends, should lead reasoners to assume that the assertors are truthful, whereas other contexts, such as rival companies advertising their products, should lead reasoners to assume that the assertors are not truthful. In a context that establishes truth-telling, reasoners should find it more difficult to make suppositions of falsehood; and, likewise, in a context that establishes lack of truth-telling, reasoners should find it more difficult to make suppositions of truth-telling. The experiment accordingly used these two contexts and a neutral control context. They are shown here with a sample inference:

1. A context to induce truth-telling:

   Two close friends describe the characteristics of their personalities:
   Jill says: I am efficient and Fay is not efficient.
   Fay says: Jill is not efficient.

2. A context to induce lack of truth-telling:

   Two business rivals advertise the characteristics of their products:
   NPE reports: NPE's product is efficient and AFC's product is not efficient.
   AFC reports: NPE's product is not efficient.

3. A neutral control context:

   Two computers describe the characteristics of their programs:
   J46 prints: J46's program is efficient and A13's program is not efficient.
   A13 prints: J46's program is not efficient.

These contexts differ in several respects (e.g. personal versus public communication, and individual versus multiple agencies), and so we carried out a test (see further on) that verified that the contexts elicited expectations of truth-telling or lying. We predicted that in the truth-inducing context, reasoners would expect that both assertors are telling the truth, and so they should find it difficult to deal objectively with the possibility that both of them are lying, or even with the possibility that one of them is lying. We predicted that in the lie-inducing context, reasoners would expect that both assertors are telling lies, and so they should find it difficult to deal objectively with the possibility that both of them are telling the truth, or even with the possibility that one of them is telling the truth.

Method

Materials. The materials were similar to those used in the factual conditions of Experiment 1. There were 16 sorts of problems: each problem contained two premises, and each premise was an assertion made by one of two speakers. The first premise contained A's assertion, which consisted of two clauses about A and B connected by a conjunction (A and B) or an inclusive disjunction (A or B, or both), and each constituent proposition was either affirmative or negative (A, not-A, B, not-B). The second premise contained B's assertion, which consisted of a single clause about A that was either affirmative or negative (A, not-A). Each of these 16 problems was presented with four suppositions for the subjects to evaluate as being consistent or inconsistent with the premises.

We constructed three versions of the problems, which differed in the contextual information given with them. The truth-inducing context was: "*two close friends describe the characteristics of their personalities*"; the lie-inducing context was: "*two business rivals advertise the characteristics of their products*"; and the control context was: "*two computers describe the characteristics of their programs*". The content of the assertions was factual and based on the following 16 adjectives: *efficient, trusted, good, smart, unique, reliable, helpful, popular, imaginative, elegant, strong, graceful, respected, refined, sensible*, and *modern*. The negation of the adjectives was explicitly based on "*not*" (e.g "*not smart*"). The 16 adjectives were randomly assigned to the 16 problems in two different ways, and the two resulting sets of materials were randomly assigned to an equal number of subjects. The name of the friends, companies, and computers were also randomly assigned to the 16 problems in two different ways. The friends were identified by 32 single-syllable same-gender names (e.g. "*Jill says . . . , Fay says . . .* "), the companies by 32 three-letter trigrams (e.g. "*DCO reports . . . , ABN reports . . .* "), and the computers by one-letter–two-digit combinations (e.g. "*P81 prints . . . , S34 prints . . .* "). The full set of problems is presented in Appendix B (along with the numbers of correct responses).

Materials Test. We presented the three sentences establishing the different contexts to a group of 15 subjects, and the order was selected randomly for each subject. The 8 men and 7 women, whose ages ranged from 21 to 26 years of age, were undergraduate students and members of the subject panel from the University of Wales College of Cardiff. They had not previously participated in an experiment on reasoning. They were asked to judge whether the assertors were more likely to tell the truth or to lie, and they responded by circling one or more of three options: (1) both individuals asserting the truth, (2) both individuals asserting a lie, and (3) one individual asserting the truth and one asserting a lie. For the truth-inducing context, 67% of subjects circled only the first option, whereas 7% circled this option for the lie-inducing context, Wilcoxon's $t = 6$, $n = 11$, $p < 0.02$. In contrast, for the lie-inducing context, 73% of subjects circled only the second option, whereas 13% circled this option for the truth-inducing context (each subject followed this pattern, apart from 6 ties, Binomial test, $n = 9$, $p = 0.5^9$). The remaining subjects for each scenario chose both the first and second option (7%) or all three options (13%). These results accordingly verify the difference between the two experimental scenarios. The neutral control context (concerning the computers) elicited 73% of identifications as truth-inducing and 7% as lying; 20% of subjects circled all three options. This result is perhaps unsurprising, given the lack of a suitably neutral category. However, we modified our use of the term "*lie*" in the main experiment accordingly and asked subjects instead whether the individuals in each of the three scenarios were asserting a truth or asserting a falsehood.

Design. Three independent groups of subjects made the four suppositional inferences with each of the 16 problems: one group with the truth-inducing context, one group with the lie-inducing context, and one group with the control context. Each subject was randomly assigned to one of the three groups and carried out the problems in a different random order.

Procedure. The subjects were tested in small groups. They were given similar instructions to those of the first experiment. They were told that they would be given assertions made by two individuals, which, depending on their group, were close friends, rival companies, or computers. Their task was to evaluate four possible states of affairs (the suppositions) as consistent or inconsistent with the premises. These four possibilities were:

A is asserting a truth and B is asserting a truth.
A is asserting a truth and B is asserting a falsehood.
A is asserting a falsehood and B is asserting a truth.
A is asserting a falsehood and B is asserting a falsehood.

They were given a practice problem with a context appropriate to their group. Each problem was presented on a separate sheet of paper, with the four suppositions listed beneath the premises. The subjects made their responses by placing a tick either under the heading "consistent" or else under the heading "inconsistent" beside each supposition. They were asked to work at their own pace, to answer the problems in the order that they were given, and not to return to a problem once they had completed it.

Subjects. Thirty subjects (14 women and 16 men), undergraduate students and members of the subject panel from the University of Wales College of Cardiff, were paid £3 per hour for their participation in the experiment. The subjects, whose ages ranged from 18 to 34, had no formal training in logic and had not previously participated in an experiment on reasoning.

Results

Table 5 summarizes the results, which are presented in detail in Appendix B. An analysis of variance (with the between-subjects factor of context and the within-subject factor of supposition type) showed that the three groups did not yield reliable differences in correct conclusions: truth-inducing (66% correct responses), lie-inducing (70% correct responses), and neutral (74% correct responses), $F(2, 27) = 0.7$, $p = 0.48$. However, as Table 5 shows, the four sorts of suppositions did yield a reliable difference in correct conclusions: A and B asserting a truth (79%); A asserting a truth and B asserting a falsehood (76%); A asserting a falsehood and B asserting a truth (67%); and A and B both asserting a falsehood (59%), $F(3, 81) = 18.95$, $p = 0.000$.

There was a marginally significant interaction between the two variables, $F(6, 81) = 1.94$, $p = 0.085$, and we carried out simple-effects analyses on it (see Winer, 1971, for the appropriateness of such planned comparisons). This analysis showed that the three groups differed reliably only in the accuracy of inferences from the supposition that A and B are both asserting a falsehood, $F(2, 108) = 3.78$, $p = 0.03$. As we predicted, reasoning from the supposition that both assertors were lying was harder in the truth-inducing context (46% correct responses) than in either the lie-inducing context (64%), Newman Keuls $Q = 4.44$, $p = 0.002$, or the neutral context (67%), $Q = 5.33$ $p = 0.003$. Contrary to our prediction, however, reasoning from the supposition that both assertors

TABLE 5
The Percentage of Correct Inferences in the Conditions of Experiment 2

	Friends	*Companies*	*Computers*	*Total*
A, B	79	79	80	*79*
A, not-B	74	72	81	*76*
not-A, B	66	66	68	*67*
not-A, not-B	46	64	67	*59*
total	*66*	*70*	*74*	

Note: A, B — A is asserting a truth and B is asserting a truth
A, not-B — A is asserting a truth and B is asserting a falsehood
not-A, B — A is asserting a falsehood and B is asserting a truth
not-A, not-B — A is asserting a falsehood and B is asserting a falsehood

The subjects' task was to judge each of these options as consistent or inconsistent with the premises.

were telling the truth was not reliably more difficult in the lie-inducing context (79% correct responses) than in either the truth-inducing context (79%), $Q = 0.00$, $p = 1.0$, or the neutral context (80%), $Q = 0.35$, $p = 0.97$. Similarly, context had no reliable effects on reasoning from the supposition that the first assertor is telling the truth and the second assertor is not telling the truth (74% correct in the truth-inducing context, and 72% correct in the lie-inducing context, in comparison with 81% in the neutral context, $Q = 1.96$, $p = 0.64$, and $Q = 2.43$, $p = 0.5$, respectively). And it had no reliable effects on reasoning from the supposition that the first assertor is not telling the truth and the second assertor is telling the truth (66% correct in the truth-inducing context, and 66% in the lie-inducing context, in comparison with 68% in the neutral context, $Q = 0.53$, $p = 0.98$, and $Q = 0.35$, $p = 0.96$, respectively).

An analysis of the data in Appendix B provided some unexpected support for the model theory. The most difficult inferences are shown there in a bold font, and a cursory examination of them reveals a systematic pattern. The difficult cases all occur when the correct response is that the supposition is consistent with the premises; and for all 16 problems in all three contexts, the condition with results in bold never yielded a higher total of correct responses than either of the other two suppositions that call for the *consistent* response, and there were just two ties (Binomial, $p = 0.33^{46}$). The pattern of problems yielding this massively significant result is at first sight difficult to interpret. However, the program implementing the model theory shows at once what causes the greater difficulty. Whenever a supposition includes the proposition that the first individual is lying, and this individual's assertion is a conjunction, then reasoners must form the *disjunctive* models corresponding to the negation of a conjunction, i.e. three models corresponding to the alternative possibilities. As an example, consider the following problem:

A asserts that c and d.
B asserts that c.

where c and d denote two factual assertions. If the supposition to be evaluated is:

A is asserting a falsehood and B is asserting a truth.

then reasoners must negate the conjunction: c and d. The result is the following disjunctive set of models:

```
 c    ¬d
¬c     d
¬c    ¬d
```

The second constituent of the supposition is that B is telling the truth, and this yields the model of B's assertion:

```
c
```

This model is consistent with the disjunctive alternatives, but it holds in only *one* of them. In contrast, when the supposition is that:

A is asserting a falsehood and B is asserting a falsehood.

it is still necessary to form the disjunctive models, but the supposition that B is asserting a falsehood yields the model:

−c

and this model is consistent with *two* of the disjunctive models. The more models are consistent with another model, the easier it should be to judge consistency, because reasoners need only find a single match. Hence, it should be easier to judge that the supposition is consistent with the premises in this second case. What causes the greatest difficulty is thus a combination of two factors: (1) the supposition about the first individual yields a set of *disjunctive* models, and (2) the supposition about the second individual yields a model that is consistent with only *one* of the models in this disjunctive set. This account is corroborated by the results for the problems where the first individual's assertion contains a disjunction, for example:

A asserts that c or d.
B asserts that c.

In these cases, the most difficult problems are those where the supposition is that A is telling the truth and the consequences of the supposition about B yield a model that is consistent with only one of the models in A's disjunctive assertion. Theories based on formal rules are unlikely to offer any account of this pattern of results, because they contain nothing equivalent to models and so cannot account for the fact that the fewer models are consistent with another model, the more difficult it should be to judge their consistency.

Discussion

The experiment showed that the context in which reasoners made a suppositional inference about matters of fact affected the accuracy with which they made it. In a truth-inducing context, they had difficulty in making inferences from the supposition that both individuals were telling a lie. This task was reliably easier in a lie-inducing context or neutral context. Presumably, the context in the first case leads the subjects to assume that the two friends are telling the truth, and this expectation makes it difficult to entertain the supposition or to follow it up inferentially. However, the subjects did not have the parallel and predicted difficulty in a lie-inducing context: They could readily cope with the supposition that the rival companies were both telling the truth. The test of the materials showed that this context genuinely sets up an expectation that both companies will lie, and so we have no reason to suppose that the context failed to create this expectaton. It is possible that the subjects were prepared to give the speakers the "benefit of the doubt"

(see Grice, 1975, for suggestions on putative maxims of conversation that lead to such assumptions). But, as the supposition of truth-telling leads to much more straightforward inferences (see also the results of Experiment 1), perhaps context exerts a marked effect only in cases where reasoners have to assume that assertions are false.

An analysis of the problems that were of greatest difficulty provided an unexpected corroboration of the model theory. Whenever it is necessary to construct a disjunctive set of models (in following up the first individual's assertion or its negation), the task of judging their consistency with another model (in following up the second individual's assertion or its negation) is difficult if this second model matches only one model in the disjunctive set.

GENERAL DISCUSSION

Truthteller–liar problems are a complex and peculiar sort of logical puzzle. Their oddity does not arise because they are about unusual individuals, who either always tell the truth or always lie. This claim is bolstered by the fact that in Experiment 1 they did not differ reliably in difficulty whether they were about such unusual individuals or about more normal individuals, who sometimes tell the truth and sometimes lie. The true source of their oddity is that their premises contain assertions that the individuals make about the truth or falsity of their own remarks. This recursive characteristic is more than an oddity, because it can lead to deep semantic paradoxes (see, e.g. Tarski, 1944). It also seems to confuse logically untrained subjects, who may lose track of the distinction between the truth-status of an individual and the content of the individual's claim. They do not always appear to grasp that a conjunction may be false and yet contain one conjunct that is true. And they do not appear to enter the psychological laboratory with a single ready-made strategy for solving *truthteller–liar* problems. They have to develop their own specialized strategies to deal with the puzzles (see Byrne & Handley, submitted; Johnson-Laird & Byrne, 1990; pace Rips, 1989, 1990). Not surprisingly, they have considerable difficulty in solving them correctly.

One strategy that logically untrained individuals do adopt is to reason by supposition—that is, to start by assuming that a particular individual in a puzzle is, say, a truthteller, and then to follow up the inferential consequences of this assumption. The strategy of making suppositions is certainly commonplace in daily life, and people make suppositions about the veracity of witnesses and others as well as suppositions about factual matters, such as the state of the weather or the economy. The new experimental paradigm that we developed enabled us both to simplify *truthteller–liar* problems and to focus on the suppositional strategy. In this paradigm, the subjects are given an explicit supposition, and, instead of having to assign a specific status to each individual in the puzzle, they merely have to decide whether or not the supposition is consistent with the premises. In other words, their task is to decide whether the supposition together with the premises is *satisfiable* in the logical sense (Tarski, 1944). A valid deduction yields a conclusion that must be true—that is, its negation is not satisfiable in any state of affairs that satisfies the premises. Hence, subjects in our task establish the separate components

that must be combined in order to identify the status of all the individuals in the puzzle.

Our failure in Experiment 1 to detect a difference between *truthtellers–liars* problems and *normals* problems contrasts with differences between problem isomorphs in other domains (for reviews, see Eysenck & Keane, 1990; Evans et al., 1993). However, in our experiment, the status of the assertor had no effect on the number of alternatives that the subjects needed to keep in mind in order to make each inference. Hence, we suggest that the number of alternatives to be kept in mind—the number of mental models (Johnson-Laird & Byrne, 1991)—is a major factor in the difficulty of these problems. This conjecture is supported by the finding that the content of the individuals' assertions affects the accuracy of inferences from them. To work out whether someone is telling the truth or lying is difficult when their assertions are, in turn, about whether they and others are telling the truth or lying. For example, the premise:

A asserts: I am a truthteller and B is a liar.

calls for the models:

A: A −B

where the annotation "A": represents A as telling the truth, and the model "A −B" represents the content of A's assertion. In this case, reasoners are likely to confuse the truth status of the individual with the content of the assertion. The task is easier when the individuals' assertions are about factual matters. For example, the premise:

A asserts: I live in Dublin and B lives in Paris

calls for the models:

A: A_{dublin} B_{paris}

and there is now much less danger of confusing the status of A with the content of A's assertion. The difficulty of premises about truth and falsity was greater for problems about *normals* than for problems about *truth-tellers* and *liars*. This phenomenon, too, bears out the tendency to confuse the status of individuals, which varies from one assertion to another in the case of *normals*, with the status of their remarks.

In general, certain sorts of suppositions are easier to work with than others. Thus, Experiment 2 established that it is easier to make inferences from the supposition that two individuals are asserting the truth about matters of fact than that they are both asserting falsehoods. But, as the experiment also suggested, this factor is likely to interact with the effects of context. Context certainly biases reasoners' expectations about whether or not individuals are telling the truth, and similar effects have been demonstrated in Wason's selection task in which reasoners have to test a conditional rule (see Evans et al., 1993). Experiment 2 showed that in a situation where two friends confide in each other, the

subjects tend to think of them as telling the truth. It is then difficult for the subjects to make inferences from the supposition that the two friends are both lying. However, the experiment did not establish the complementary effect: The subjects did not find the task of making inferences from the supposition that two individuals were telling the truth any more difficult in a context that predisposed them to expect the individuals to lie. This phenomenon may reflect the customary Gricean convention that people tell the truth (Grice, 1975) or the relative ease of suppositions about truth as opposed to suppositions about falsity.

Another aspect of the model theory was corroborated by Experiment 2. The theory predicts that disjunctive models are a general source of difficulty—that is, the more models reasoners have to construct, the longer the inferential task will take and the more errors it is likely to induce (see Byrne & Johnson-Laird, 1989, 1992; Johnson-Laird & Byrne, 1989; Johnson-Laird, Byrne & Tabossi, 1989). In fact, regardless of the context, the subjects in Experiment 2 were most likely to err when they had to construct a set of disjunctive models. They had to do so either because they had to follow up the supposition that a disjunctive assertion was true or else because they had to follow up the supposition that a conjunctive assertion was false. The task of negating a conjunction is independently known to cause problems (see Handley & Byrne, in preparation). What was particularly problematic about a disjunctive set of models in our paradigm was establishing a match between the model (of the other individual's assertion) and just one of the models in the disjunctive set. This difficulty was apparent in all 16 of the different sorts of problem and for all three different sorts of context: It was not controverted by a single set of responses in any of the 48 data sets.

Could our results be explained by a theory based on formal rules, such as Rips's (1989, 1990) account of *knight-and-knave* problems? Perhaps. The theory could certainly postulate a tendency to confuse the status of individuals—*knight* or *knave*—with the status of their assertions, and in this way it might account for the phenomena of Experiment 1. But one difficulty as we showed in the earlier section on the two theories, is that the inferences about matters of fact call for longer formal derivations than do inferences about matters of truth and falsity. This difference makes exactly the opposite prediction to the results of the experiment. Experiment 2 raises further difficulties for rule theories, because they have nothing that corresponds to models and no principled machinery to explain why matching one model to a disjunctive set is more difficult when there is only one corresponding model in the set as opposed to two.

In everyday thinking, you must often assess whether the information given to you could be true before you go on to reason about its consequences. Our investigation has focused on one way that you make this assessment—you make inferences from the given information in order to decide whether or not it is at least internally consistent. Such a test is a minimal one. It amounts, in our terms, to whether or not a model, or a set of models, can be constructed from the premises. Inconsistency is revealed by the inability to find any model of the premises and supposition (i.e. the computer program implementing the theory returns the null model). In daily life, you generally assess not merely consistency, but also the likelihood that the given information is true (see e.g. Kahneman, Slovic, & Tversky, 1982), but even this assessment might be made by considering the proportion of possible models that satisfy the given information.

REFERENCES

Anno, M., & Nozaki, A. (1984). *Anno's hat tricks.* London: Bodley Head.

Austin, J.L. (1970). Truth. In J.O. Urmson and J.G. Warnock (Eds.), *Philosophical papers of J.L. Austin*, 2nd Edition. Oxford: Oxford University Press.

Barwise, J., & Etchemendy, J. (1987). *The liar: an essay in truth and circularity.* New York: Oxford University Press.

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.

Byrne, R.M.J. (1989a). Suppressing valid inferences with conditionals. *Cognition*, 31, 61–83.

Byrne, R.M.J. (1989b). Everyday reasoning with conditional sequences. *Quarterly Journal of Experimental Psychology, 41A*, 141–166.

Byrne, R.M.J., & Handley, S.J. (1992). Reasoning strategies. *Irish Journal of Psychology, 13*, 111–124.

Byrne, R.M.J., & Handley, S. (1993). The nature and development of meta-deductive reasoning strategies. In K. Ryan & R.F.E. Sutcliffe (Eds.), *AI and Cognitive Science '92* (pp. 59–70). London: Springer-Verlag.

Byrne, R.M.J., & Handley, S. (submitted). *Reasoning strategies in suppositional deductions.*

Byrne, R.M.J., Handley, S. & Johnson-Laird, P.N. (1992). Advances in the psychology of reasoning: Meta-deduction. In M.T. Keane & K. Gilhooly (Eds.), *Advances in the Psychology of Thinking, Vol. 1* (pp. 127–145). London: Harvester Wheatsheaf.

Byrne, R.M.J., & Johnson-Laird, P.N. (1989). Spatial reasoning. *Journal of Memory and Language, 28*, 564–575.

Byrne, R.M.J., & Johnson-Laird, P.N. (1992). The spontaneous use of propositional connectives. *Quarterly Journal of Experimental Psychology, 44A*, 89–110.

Byrne, R.M.J., Johnson-Laird, P.N., & Handley, S. (1993). Who's telling the truth . . . Cognitive processes in meta-deductions. In H. Sorenson (Ed), *AI and Cognitive Science '91* (pp. 221–233). London: Springer-Verlag.

Dewdney, A.K. (1989). People puzzles: Theme and variations. *Scientific American, 260*, 1 (January), 88–91.

Evans, J.St.B.T. (1972). *Deductive reasoning and linguistic usage (with special reference to negation).* Unpublished PhD Thesis, University of London.

Evans, J.St.B.T. (1990). Reasoning with knights and knaves: A discussion of Rips. *Cognition, 36*, 85–90.

Evans, J.St.B.T., Newstead, S.N., & Byrne, R.M.J. (1993). *Human reasonings: The psychology of deduction.* Hove, UK: Lawrence Erlbaum Associates, Ltd.

Eysenck, M.W., & Keane, M.T. (1990). *Cognitive psychology: A student's handbook.* Hove, UK: Lawrence Erlbaum Associates, Ltd.

Fujimura, K. (1884). *The Tokyo puzzles.* New York: Scribner, 1978.

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

Goffman, E. (1959). *The presentation of self in everyday life.* New York: Doubleday.

Grice, H.P. (1975). Logic and conversation. In P. Cole and J.L. Morgan (Eds.), *Syntax and semantics. Vol. 3: speech acts.* New York: Seminar Press.

Handley, S.J., & Byrne, R.M.J. (in preparation). *The negation of conjunctions and disjunctions.* University of Plymouth, England.

Jeffrey, R. (1981). *Formal logic: Its scope and limitations*, second edition. New York: McGraw-Hill.

Johnson-Laird, P.N., & Barres, P.E. (1994). When *or* means *and*: A study in mental models. In A. Ram & K. Eiselt (Eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 475–478). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Johnson-Laird, P.N., & Byrne, R.M.J. (1989). *Only* reasoning. *Journal of Memory and Language, 28*, 313–330.

Johnson-Laird, P.N., & Byrne, R.M.J. (1990). Meta-logical puzzles: Knights, knaves and Rips. *Cognition, 36*, 69–84.

Johnson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction*. Hove, UK: Lawrence Erlbaum Associates, Ltd.
Johnson-Laird, P.N., Byrne, R.M.J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review, 99*, 418–439.
Johnson-Laird, P.N., Byrne, R.M.J., & Tabossi, P. (1989). Reasoning by model: The case of multiple quantification. *Psychological Review, 96*, 658–673.
Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty*. Cambridge: Cambridge University Press.
Keane, M.T., Ledgeway, T., & Duff, S. (1994). Constraints on analogical mapping: A comparison of three models. *Cognitive Science, 18*, 287–334.
Kotovsky, K., & Simon, H.A. (1989). What makes some problems really hard: Explorations in the problem space of difficulty. *Cognitive Psychology, 22*, 143–183.
Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy, 72*, 690–716.
Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
Newell, A., & Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
Prawitz, D. (1965). *Natural deduction: A proof-theoretical study*. Stockholm: Almqvist and Wiksell.
Quine, W.V.O. (1974). *Methods of logic (3rd edition)*. London: Routledge & Kegan Paul.
Rips, L.J. (1983). Cognitive processes in propositional reasoning. *Psychological Review, 90*, 38–71.
Rips, L.J. (1989). The psychology of knights and knaves. *Cognition, 31*, 85–116.
Rips, L.J. (1990). Paralogical reasoning: Evans, Johnson-Laird, and Byrne on liar and truth-teller puzzles. *Cognition, 36*, 291–314.
Rips, L.J. (1994). *The psychology of proof*. Cambridge: MA: MIT.
Smullyan, R.M. (1978). *What is the name of this book? The riddle of Dracula and other logical puzzles*. Englewood Cliffs, NJ: Prentice-Hall.
Tarski, A. (1944). The semantic conception of truth. *Philosophy and phenomenological research, 4*, 341–375.
Wason, P.C., & Johnson-Laird, P.N. (1972). *Psychology of reasoning: Structure and content*. London: Batsford.
Winer, B.J. (1971). *Statistical principles in experimental design*, Third edition. New York: McGraw-Hill.

Revised manuscript received 14 November 1994

APPENDIX A

A's assertion is listed across the horizontal line, and the connective it contained was either a conjunction or a disjunction, indicated in the left margin. It referred to characteristics of A and B, asserting that, for example, A possessed the characteristic, which we represent as a, or did not possess it, which we represent as −a. The subjects were presented with four suppositions:

a b	A is telling the truth and B is telling the truth
a −b	A is telling the truth and B is lying
−a b	A is lying and B is telling the truth
−a −b	A is lying and B is lying

We list each of these suppositions for each problem—the consistent possibilities are in italics, the remaining possibilities are inconsistent. The frequency correct is therefore for the judgement that a possibility is consistent when it is so, and for the judgement that it is inconsistent when it is so. The frequencies correct for the truthtellers–liars group (T–L) and for the normals group (N) are listed separately. Each subject carried out each inference once, and so the maximum number in a cell is 16 for each group. The inferences were based on truth-status, listed first, or on location, listed second.

Truth Status Assertions (for Both Truthtellers/Liars and Normals Context)

	A's assertion refers to:							
	A Truthtelling B Truthtelling		*A Truthtelling B Lying*		*A Lying B Truthtelling*		*A Lying B Lying*	
	T–L	*N*	*T–L*	*N*	*T–L*	*N*	*T–L*	*N*
A's assertion contains the connective: AND								
B's assertion refers to: A truthtelling								
Supposition								
a b	*15*	*16*	13	11	9	12	9	15
a −b	14	13	11	9	6	6	8	2
−a b	15	15	11	9	12	5	9	5
−a −b	*10*	*10*	*12*	*8*	*8*	*9*	9	9
B's assertion refers to: A lying								
Supposition								
a b	12	14	13	9	4	7	9	6
a −b	6	2	*10*	*13*	13	13	10	10
−a b	*7*	*13*	*11*	*15*	11	6	*7*	*12*
−a −b	8	8	9	8	5	7	7	6
A's assertion contains the connective: OR . . . OR BOTH								
B's assertion refers to: A truthtelling								
Supposition								
a b	*15*	*14*	*11*	*9*	*9*	*7*	7	9
a −b	11	13	12	9	11	9	9	1
−a b	13	13	12	9	11	7	8	7
−a −b	*10*	*13*	8	7	7	7	4	4
B's assertion refers to: A lying								
Supposition								
a b	12	12	10	13	5	4	8	4
a −b	*14*	*13*	*12*	*11*	10	13	*7*	*8*
−a b	7	3	*10*	*13*	11	8	7	6
−a −b	11	7	10	9	8	3	9	6

Note: T–L = truthtellers–liars group; N = normals group.

Location Assertions (for Both Truthteller/Liars and Normals Context)

	A's assertion refers to:							
	A in Dublin B in Dublin		*A in Dublin B in Paris*		*A in Paris B in Dublin*		*A in Paris B in Paris*	
	T–L	*N*	*T–L*	*N*	*T–L*	*N*	*T–L*	*N*
A's assertion contains the connective: AND								
B's assertion refers to: A in Dublin								
Supposition								
a b	*13*	*16*	*15*	*13*	16	15	16	16
a −b	15	15	16	14	*14*	*15*	*15*	*15*
−a b	*7*	*7*	*5*	*11*	*14*	*13*	*13*	*14*
−a −b	*9*	*12*	*10*	*14*	*10*	*11*	*10*	*14*
B's assertion refers to: A in Paris								
Supposition								
a b	16	10	15	15	*14*	*15*	*14*	*16*
a −b	*14*	*15*	*14*	*16*	16	14	15	15
−a b	*13*	*15*	*11*	*15*	*6*	*9*	*5*	*8*
−a −b	*11*	*13*	*10*	*13*	*11*	*14*	*12*	*13*
A's assertion contains the connective: OR . . . OR BOTH								
B's assertion refers to: A in Dublin								
Supposition								
a b	*14*	*16*	*15*	*16*	*4*	*2*	*3*	*5*
a −b	*3*	*8*	*4*	*8*	*11*	*15*	*12*	*16*
−a b	11	8	12	10	*13*	*14*	*12*	*13*
−a −b	*10*	*14*	*12*	*13*	*10*	*13*	*10*	*12*
B's assertion refers to: A in Paris								
Supposition								
a b	*5*	*5*	*2*	*4*	*13*	*16*	*13*	*16*
a −b	*10*	*16*	*12*	*14*	*6*	*5*	4	*7*
−a b	*14*	*13*	*14*	*14*	10	9	12	7
−a −b	*9*	*12*	*9*	*14*	*10*	*12*	*11*	*14*

Note: T–L = truthtellers–liars group; N = normals group.

APPENDIX B

The conventions are identical to those for Appendix A. The frequency correct for the truth-inducing group (T) is listed first, the frequency for the lie-inducing group (L) is listed second, and the frequency for the neutral group (N) is third. Ten subjects in each group solved each problem once, and hence the maximum number in a cell is 10. The most difficult problems for all three groups—that is, the problems where the frequency correct is on or below 5—are in bold. The content of the problems was based on 16 different adjectives, and only one of them, efficiency, is used here as an illustration.

	A's assertion refers to:											
	A efficient B efficient			*A efficient B not efficient*			*A not efficient B efficient*			*A not efficient B not efficient*		
	T	*L*	*N*	*T*	*L*	*N*	*T*	*L*	*N*	*T*	*L*	*N*
A's assertion contains the connective: AND												
B's assertion refers to: A efficient												
Supposition												
a b	*9*	*9*	*10*	*9*	*10*	*10*	7	10	10	10	9	9
a −b	9	10	10	9	9	10	*7*	*7*	*10*	*10*	*8*	*9*
−a b	***4***	***2***	***1***	***6***	***3***	***0***	*8*	*9*	*10*	*7*	*9*	*9*
−a −b	*5*	*9*	*9*	*6*	*8*	*9*	***3***	***5***	***0***	***4***	***3***	***2***
B's assertion refers to: A not efficient												
Supposition												
a b	10	10	10	9	10	10	*8*	*10*	*10*	*9*	*9*	*10*
a −b	*8*	*10*	*10*	*8*	*10*	*10*	10	8	10	8	9	10
−a b	*9*	*8*	*10*	*7*	*8*	*10*	***5***	***4***	***1***	***5***	***2***	***1***
−a −b	***3***	***3***	***1***	***3***	***3***	***2***	*6*	*7*	*10*	*5*	*8*	*10*
A's assertion contains the connective: OR . . . OR BOTH												
B's assertion refers to: A efficient												
Supposition												
a b	*10*	*10*	*10*	*10*	*9*	*10*	***5***	***4***	***3***	***4***	***2***	***3***
a −b	***3***	***3***	***3***	***5***	***2***	***4***	*6*	*8*	*9*	*9*	*9*	*10*
−a b	7	8	7	8	7	6	*6*	*8*	*9*	*8*	*10*	*10*
−a −b	*6*	*8*	*8*	*5*	*7*	*10*	4	8	7	6	7	7
B's assertion refers to: A not efficient												
Supposition												
a b	***5***	***2***	***4***	***3***	***4***	***3***	*9*	*9*	*10*	*9*	*9*	*10*
a −b	*9*	*9*	*10*	*9*	*8*	*9*	***4***	***4***	***3***	***3***	***1***	***2***
−a b	*7*	*8*	*9*	*7*	*8*	*9*	5	6	8	5	6	8
−a −b	6	6	7	5	5	8	*5*	*7*	*9*	*6*	*8*	*8*

Note: T = truth-inducing group; L = lie-inducing group; N = neutral group.