Memory & Cognition
2004, 32 (1), 96-106

Co-reference and reasoning

CLARE R. WALSH
Brown University, Providence, Rhode Island
and
P. N. JOHNSON-LAIRD
Princeton University, Princeton, New Jersey

Co-reference occurs when two or more noun phrases refer to the same individual, as in the following inferential problem: Mark is kneeling by the fire or he is looking at the TV but not both. / Mark is kneeling by the fire. / Is he looking at the TV? In three experiments, we compared co-referential reasoning problems with problems referring to different individuals. Experiment 1 showed that co-reference improves accuracy. In Experiment 2, we replicated that finding and showed that co-reference speeds up both reading and inference. Experiment 3 showed that the effects of co-reference are greatest when the premises and the conclusion share co-referents. These effects led the participants to make illusory inferences—that is, to draw systematically invalid conclusions. The results are discussed in terms of the mental model theory of reasoning.

Deductive reasoning occurs in a variety of domains in science and everyday life. It yields a conclusion from a set of premises, and such an inference is valid if the conclusion must be true given the truth of the premises. The present article concerns *sentential reasoning*—that is, deductions that hinge on negation and sentential connectives, such as *if*, *and*, and *or*. Consider, for instance, the following inference:

Rachel is climbing up the stairs or she is cooking at the stove but not both.

Rachel is not climbing up the stairs.

Is she cooking at the stove? [Answer: yes.]

The inference is valid, because if its premises are true then its conclusion must be true too. This is an example of an inference with a co-referential actor, because readers understand that *Rachel* and *she* refer to the same person. What effect does such a co-reference have on deductive reasoning? No definite answer is to be found in the psychological literature, and the present article aims to remedy this deficiency.

If individuals use formal rules of inference in reasoning, then they should possess the following rule for disjunctions:

A or B.

Not-A.

Therefore, B.

They can make the inference above by applying this rule to the logical form of the premises, setting *A* equal to *Rachel is climbing up the stairs* and *B* equal to *Rachel is cooking at the stove* (see, e.g., Braine & O'Brien, 1998; Rips, 1994). The use of the rule should be the same whether or not clauses *A* and *B* are co-referential.

In contrast, the mental model theory postulates that individuals with no training in logic use the meaning of assertions and their general knowledge to construct mental models of the possibilities that are compatible with the premises (Johnson-Laird & Byrne, 1991). Each mental model represents a possibility. A conclusion is judged to be valid if it holds in all the mental models of the premises. Conversely, it is judged as invalid (i.e., not necessarily the case) if reasoners find a counterexample—that is, a model of the premises in which the conclusion is false.

Because working memory is limited, mental models are governed by the *principle of truth:* by default, they represent only true possibilities and, within them, the clauses in the premises only when they are true (Johnson-Laird & Byrne, 2002). If a clause is false in a possibility, then it will not be represented in a mental model of the possibility. For instance, given an *exclusive* disjunction of the form *A or B but not both*, several experiments have shown that individuals list as possible just the following

At the time most of this research was conducted, the first author was supported by an Enterprise Ireland PhD fellowship, a Government of Ireland Scholarship from the Council for Humanities and Social Sciences, a Dublin University Postgraduate Award, and a postdoctoral fellowship from the Educational Testing Service. The second author was supported by Grant BCS 0076287 from the National Science Foundation for investigation of strategies in reasoning. Some of the results were presented at the 23rd meeting of the Cognitive Science Conference in Edinburgh in August, 2001. We thank Ruth Byrne, Sam Glucksberg, Uri Hasson, Markus Knauff, Mike Oaksford, Yingrui Yang, Lauren Ziskind, and two anonymous reviewers for their helpful comments on this research. Correspondence concerning this article should be addressed to C. Walsh, Department of Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912 (e-mail: clare_walsh@brown.edu).

two cases (see, e.g., Barres & Johnson-Laird, 2003; Johnson-Laird & Savary, 1996):

A
B

According to the theory, individuals make a mental note of what is false in each of these possibilities. If they retain these notes, then they can in principle use them to flesh out the mental models into *fully explicit* models:

A ¬ B
¬ A B.

where "¬" denotes negation. Otherwise, they will be vulnerable to various sorts of illusory inferences that seem compelling but are in fact invalid (see, e.g., Johnson-Laird & Savary, 1996). We will return to these illusions in the account of Experiment 3.

To make an inference of the sort

A or B but not both.
A.
Therefore, not-B,

reasoners can match the categorical information in the second premise with the first of the models above and then flesh out the model to draw the conclusion *not-B*. In contrast, to make an inference of this sort with a negative categorical premise, such as

A or B but not both.
Not-A.
Therefore, B,

reasoners must use the categorical information to eliminate a model—the first in the set above—and must find the second model and assert the information that it contains. In general, the process of matching a premise to a model is easier than that of mismatching a premise to a model (i.e., using a premise to negate a model). Hence, the theory predicts that the first sort of inference should be easier than the second sort, and the results of previous experiments have corroborated this prediction (see, e.g., Johnson-Laird & Byrne, 1991).

A biconditional of the form *If and only if not-A then B* has exactly the same fully explicit models as the preceding disjunction does. Most people, however, do not immediately realize the equivalence. They consider instead the mental models of the biconditional:

¬ A B
. . .

The first model represents the salient possibility in which the antecedent, *not-A*, and the consequent, *B*, are both true. The second model, denoted by the ellipsis, is a placeholder with no explicit content. It represents the possibilities in which the antecedent and the consequent are both false. If individuals retain the mental note of this information, they can use it to construct fully explicit models of the biconditional, which are the same as those for the exclusive disjunction *A or B but not both*.

The meaning of a premise, co-reference, and background knowledge can all modulate the basic meanings of sentential connectives (Johnson-Laird & Byrne, 2002). Modulation occurs, for example, when knowledge prevents the construction of a mental model. Thus, a disjunction such as

Rachel ate salmon for dinner or she ate fish

would ordinarily be compatible with the possibility in which she ate salmon but not fish. However, this possibility is ruled out by the co-reference of *Rachel* and *she* and by the knowledge that salmon is fish.

The recovery and representation of co-reference is central to comprehension. Indeed, the ease of establishing co-referential relations enhances the understanding and memory of discourse (see, e.g., Garnham, Oakhill, & Johnson-Laird, 1982). A noun phrase can lead to the introduction of a new entity into a mental model of discourse, and subsequent references yield the representation of new properties of that entity and new relations into which it enters (see Johnson-Laird, 1983, chap. 14). It follows that a description containing a co-reference yields mental models containing fewer distinct referents than would a corresponding passage referring, instead, to different individuals. Other theories of discourse make similar predictions. Descriptions of events can be integrated more easily if they share referents with existing mental structures (Gernsbacher, 1990) or situation models (Zwaan & Radvansky, 1998). It follows that both the comprehension of premises and the deduction of conclusions should be easier when there is co-reference than when there is not (see also Shastri & Ajjanagadde, 1993). An alternative possibility is that co-reference may lead to a fan effect. The more facts are learned about a concept, the longer it takes to recall any one of them. If the same information is stored in more than one model, there may be some interference during retrieval (Zwaan & Radvansky, 1998). Hence, deductive inference may be easier when information in the different models is distinct. In a pioneering study, Bouquet and Warglien (1999) showed that reasoning from disjunctive premises yielded a greater number of valid inferences when the clauses were co-referential than when they were not (although it is unclear whether the difference was statistically significant). In the three experiments presented here, we examined a wider variety of sorts of inference and sorts of co-reference.

The model theory postulates that a model has the same structure as the situations that they represent (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991). Consider, for instance, a disjunction with a co-referential actor, such as

Rachel is climbing up the stairs or she is cooking at the stove but not both.

The mental models of this disjunction represent three referents (*Rachel*, *the stairs*, and *the stove*), the relations

between *Rachel* and the other two referents, and the fact that they hold in two alternative possibilities. For simplicity, rather than spell out the complete relational structure in a set of models, we use the following diagram to denote the two alternative models:

Rachel: climbing stairs
cooking at stove

A colon indicates that the preceding item denotes a referent common to each of the following models. In this case, the single referent followed by a colon denotes Rachel as the actor common to both actions. The mental models of a disjunction with two different actors, such as

Rachel is climbing up the stairs or David is cooking at the stove but not both,

call for an additional token to represent *David*:

Rachel: climbing stairs
David: cooking at stove

The following disjunction has a co-referential action:

Rachel is climbing up the stairs or David is climbing up the stairs but not both.

The mental model theory makes no predictions about how people mentally represent a sentence with shared predicates. However, such a sentence could be represented more concisely than a sentence with different predicates.

The mental models of the disjunction with one actor are simpler than those of the disjunction with two actors. Hence, the theory predicts that inferences based on co-referential actors should be easier than those based on two distinct actors. The case of co-referential actions is less clear-cut, but since there are only three referents, inferences should be easier than those based on two actors carrying out distinct actions.

In summary, there are two main mechanisms through which co-reference may improve reasoning. First, for some co-referential sentences, knowledge may prevent the construction of models that are inconsistent with the premises. Second, sentences with co-reference may have an initial representation that is more concise. Hence, the working memory load may be reduced.

EXPERIMENT 1

In this experiment, we examined inferences based on biconditionals of the form *If and only if not-A then B* and on disjunctions of the form *A or B but not both*. As we illustrated earlier, these two sorts of assertion are logically equivalent, as their fully explicit models show, but they have different mental models. There were eight sorts of inferences. Four were based on disjunctions:

1. Either A or B but not both.

   A.

   What follows?

2. Either A or B but not both.

   Not-A.

   What follows?

3. Either A or B but not both.

   B.

   What follows?

4. Either A or B but not both.

   Not-B.

   What follows?

The remaining four were inferences based on equivalent biconditionals.

In two main sorts of sentences (i.e., those with one actor and those with two actors), co-reference was manipulated. We carried out a norming study to test whether it was plausible for the actions in the biconditionals and disjunctions to occur simultaneously in such sentences. The participants in the norming study were 24 undergraduate students at Princeton University. They read sentences consisting of conjunctions of events and judged each sentence in a binary way according to whether it was possible for the two actions to occur together. Hence, for a *one-actor sentence*, such as

Sarah is sitting in the armchair or Sarah is opening the front door but not both,

the participants judged the following conjunction:

Sarah is sitting in the armchair and Sarah is opening the front door.

The *two-actor sentences* varied in terms of whether the actors carried out the same actions or not, as is exemplified in the following:

Brian is standing by the fireplace or Joanne is looking in the mirror but not both.

Linda is looking out the window or Richard is looking out the window but not both.

Graham is standing on the scales or Carol is standing on the scales but not both.

The participants judged conjunctions in the same way for each of the biconditional and disjunctive premises in Experiment 1. The one-actor conjunctions were judged to have low plausibility. They were rated as possible in only 35% of the cases, whereas the two-actor conjunctions were rated as possible in 95% of the cases (Wilcoxon test, $z = 4.30, p < .001$). As we had surmised, the third sort of two-actor conjunctions, in which it was less plausible for both individuals to carry out the same action simultaneously, were judged as possible slightly less often (86%) than the two other sorts, in which it was highly plausible that the two actions could occur simultaneously (both 100%; Wilcoxon test, $z = 2.59, p < .05$). Since the difference was quite small, we will collapse the three sorts of two-actor sentences henceforth.

Table 1
Percentages of Correct Responses to the Eight Forms of Inference in Experiment 1

Compound Premise	Mental Model	Categorical Premise A	Not-A	B	Not-B
If and only if not-A then B.	¬ A B	78	**91**	**94**	46
A or B but not both.	A . . . B	**94**	68	**94**	68

Note—Inferences are in bold when the categorical premise matches a mental model of the first premise.

As we showed in the introduction, the model theory predicts that inferences should be easier when the categorical premise matches information in a mental model of the *compound* premise (i.e., the biconditional or disjunction) than when the categorical premise mismatches information in a mental model of the compound premise. The main aim of the experiment, however, was to examine the effects of co-reference.

Method

Participants. We tested individually 30 undergraduates (15 men and 15 women) from Princeton University in return for course credit. The participants ranged in age from 18 to 23 years, with a mean age of 20.

Design and Materials. The first premise in each inference was a biconditional (*If and only if not-P then Q*) or an exclusive disjunction (*P or Q but not both*). The second premise was a categorical assertion or a categorical denial of either the first or the second clause in the first premise (*P*, *not-P*, *Q*, *not-Q*). Accordingly, there were eight forms of inference, and each of them yielded a valid conclusion. Each form of inference occurred with the two main sorts of sentence (that with one actor or that with two actors) in the referential manipulation. We used the three sorts of two-actor sentences above, and so there were a total of 32 problems (8 one-actor problems and 24 two-actor problems). In every case, the compound premise contained the same number of words. A full set of materials in the form of exclusive disjunctions is given in the Appendix. There were also 4 filler problems based on inclusive disjunctions, for which the correct response was that nothing followed from the premises—that is, they did not yield a valid conclusion. Each participant acted as his or her own control and carried out the 36 inferences in a different random order. The materials were devised from a set of common actions, as was illustrated earlier, and a list of common two-syllable male and female names.

Procedure. The participants were tested individually in a quiet room. They were told that the experiment concerned reasoning and that they would be asked to judge what conclusion, if any, could be drawn from pairs of sentences. They were not told that their responses were being timed. The problems were presented on a computer screen under the control of a computer program. The participants read the on-screen instructions. The key instructions were as follows:

> This program is going to present pairs of sentences on the screen. After each pair of sentences it will ask you whether you can draw a conclusion from them. Please read the sentences carefully and type your answer on the screen. When you have finished typing your response, please press the return key.

They then carried out 5 practice problems followed by the 36 problems in the experiment. The practice problems were deductive reasoning problems of a different type than the experimental problems. These problems allowed the participants to familiarize themselves with the format of the problems and with how to respond, but the participants were not given feedback on their responses. For each problem, the two premises were presented together on the screen, along with the question:

> What conclusion, if any, can you draw?

The participants responded by typing their answers, and their latencies were measured from the presentation of the premises to the first keypress.

Results and Discussion

Categorical assertions or denials (depending on the problem) of the clause not referred to in the second premise were scored as correct. In Table 1, the percentages of correct responses to the eight forms of inference (collapsing over their contents) are presented. As the table shows, there was no reliable difference between reasoning from biconditionals (77% correct) and reasoning from disjunctions (81% correct). Biconditionals normally have an advantage because they have only one mental model with an explicit content, but this advantage was probably offset by their negated antecedents. As the model theory predicts, when the categorical premise matched an event that was explicitly represented in a mental model of the compound premise (93% correct), the participants were more accurate than when it did not (65%; Wilcoxon test, $z = 4.29, p < .0001$). This effect could merely reflect the surface matching of clauses in the premises. One result, however, is more readily explained in terms of models. When the categorical premise (*not-B*) negated the second clause of the compound premise, the participants were more accurate in reasoning from disjunctions (68% correct) than from biconditionals (46%; Wilcoxon test, $z = 2.77, p < .006$). *Not-B* does not match the mental models of the disjunctions or the biconditionals. However, reasoners should find it easier to flesh out the disjunctions, which already have two mental models with explicit contents, than to flesh out the biconditionals, which have only one mental model with an explicit content (see Schaeken, Johnson-Laird, Byrne, & d'Ydewalle, 1995).

Co-reference did not interact reliably with the forms of inference. Likewise, there were no reliable differences among the three sorts of two-actor sentences in either percentages of correct responses or their latencies. Hence, Table 2 shows the percentages of correct responses for one-actor and two-actor inferences. As the theory predicted, the participants tended to make a slightly greater percentage of accurate inferences from the one-actor problems (82%) than from the two-actor problems (78%; Wilcoxon test, $z = 1.83, p < .04$). Hence, there is evidence replicating Bouquet and Warglien's (1999) finding. There were no significant differences between the latencies of the correct responses to the different sorts of

Table 2
Percentages of Correct Responses for the Four Sorts of Referential Content in Experiment 1

Inferential Content	Biconditional	Disjunction	Mean
One actor	78	85	82
Two actors	77	80	78

inference. Co-reference had no significant effect on the latencies of correct responses.

Co-reference can enhance reasoning. The fact that individuals are more accurate in making inferences from premises with one co-referential actor is consistent with Bouquet and Warglien (1999), but it holds in the present study for both disjunctive and biconditional inferences. This result may be attributed in part to the pragmatic effects of knowledge. The one-actor premises concerned two actions that cannot occur simultaneously. This factor should assist performance for affirmative inferences by eliminating the possibility that both actions occurred. However, pragmatic effects alone seem unlikely to account for the result, because the improvement occurred for inferences based on both affirmative and negative categorical premises. What happens when there is no pragmatic effect for one-actor inferences? In Experiment 2, we aimed to answer this question.

EXPERIMENT 2

In this experiment, we examined separately the effects of co-reference on the times taken to read the premises and on reasoning. Only disjunctive premises were used because no reliable differences in referential effects between disjunctions and biconditionals were detected in the previous experiment. In all conditions, the two actions could occur simultaneously (as was corroborated in the norming study reported earlier). The computer-controlled procedure yielded reading times for the disjunctive and categorical premises and the times to respond to the inferential questions.

Method

Participants. Thirty undergraduates (8 men and 22 women, mean age 21 years) from the University of Dublin were individually tested in return for course credit.

Design and Materials. The participants carried out four forms of inference. The first premise was an exclusive disjunction. The second premise was a categorical assertion or a categorical denial of either the first or the second clause in the disjunction. Finally, there was an affirmative question as to whether the other clause in the disjunction followed validly. A typical problem was as follows:

Karl is eating at the table or Sue is eating at the table but not both.

Sue is not eating at the table.

Is Karl eating at the table?

The disjunctive premises were of three sorts: (1) one-actor premises (e.g., *Mark is kneeling by the fire or he is looking at the TV but not both*), (2) one-action premises (e.g., *Karl is eating at the table or Sue is eating at the table but not both*), and (3) no-co-reference premises (e.g., *Paul is standing by the fireplace or Dave is looking in the mirror but not both*).

As the examples illustrate, the actions in all three sorts of disjunction could be performed simultaneously, and all three sorts of disjunction contained the same number of words. We selected a sample of four one-actor premises and four one-action premises, and we obtained independent ratings of whether the two actions in the premises could occur simultaneously. The ratings were done by the same 24 undergraduate students who had taken part in the norming study in Experiment 1, and the task was identical. The two actions were rated as simultaneously possible in 100% of the one-actor problems and in 96% of the one-action problems. The three types of disjunction were presented with each of the four forms of inference. There were also eight filler items, and so each participant carried out a total of 20 problems. The problems were presented in a different random order to each participant. We devised a set of common one-syllable names paired with everyday actions, and those lexical contents were rotated so that with appropriate referential adjustments they were presented equally often with the three sorts of disjunction in the experiment as a whole.

Procedure. The problems were presented on a computer screen using the E-Prime package. The participants read the on-screen instructions. The key instructions were as follows:

During the experiment, you will be presented with sentences on the screen. When you have read each sentence you should press the space-bar to continue. When you have finished reading two sentences, you will be presented with a question. Please read the sentences and the question carefully. You should answer the question by pressing the "Yes," "No," or "Cannot tell" key on the keyboard.

The "yes," "no," and "cannot tell" keys corresponded to the T, O, and U keys, respectively. The participants were not told that their responses would be timed. They then completed 5 practice problems, which were of a different form than the experimental problems, followed by the set of 20 problems. For each problem, the premises and questions were presented one by one on the screen. The premises remained on the screen until the participants had pressed one of the three keys to answer the question. The program recorded separately the times that it took the participants to read each of the premises and to answer the question in each problem.

Results and Discussion

In Table 3, the percentages of correct responses for the four forms of inference are presented. When the categorical premise matched an event explicitly represented in a mental model of the disjunctive premise, the participants were more accurate (94%) than when it did not (77%; Wilcoxon test, $z = 2.36, p < .02$). They were more accurate when the categorical premise referred to the first clause of the disjunction (88%) than when it referred to the second (83%; Wilcoxon test, $z = 2.18, p < .03$). There was a marginal interaction, showing that the latter difference occurred only when the categorical premise did not match a mental model (Wilcoxon test, $z = 1.88, p < .06$).

Table 4 shows the percentage of correct responses and the reading and response times for each of the three sorts of referential sentences for those problems to which the participants responded correctly. The participants were more accurate in making inferences from one-actor problems (90%) than from no-co-reference problems (83%) and from one-action problems (83%; Wilcoxon tests, $z = 1.81, p < .04$, one-tailed, and $z = 1.90, p < .06$, two-tailed, respectively). There was no reliable difference between one-action and no-co-reference problems.

The reading and response times for correct responses show a consistent pattern: The participants read and solved the one-action problems faster than the one-actor problems and the one-actor problems faster than the no-

Table 3
Percentages of Correct Responses to the Four Forms of Inference Based on Exclusive Disjunctions in Experiment 2

	Categorical Premise			
Disjunction	A	Not-A	B	Not-B
A or B but not both	93	82	94	71

Table 4
Percentages of Correct Responses and Reading and Response Times for Correctly Answered Problems in Experiment 2

Referential Content	% Correct Responses	Time (Sec)			
		Reading of Disjunctive Premise	Reading of Categorical Premise	Response	Total
One actor	90	5.2	3.1	3.1	11.4
One action	83	4.6	2.6	2.4	9.6
No co-reference	83	6.6	4.3	4.0	14.9

co-reference problems. This trend was reliable for reading of the disjunctive premises (Page's L test, $z = 3.87, p < .00007$), for reading of the categorical premises (Page's L test, $z = 4.13, p < .00003$), and for responding to the inferential questions (Page's L test, $z = 3.23, p < .0007$).

As in the previous experiment, a single co-referential actor improved the accuracy of reasoning. The actor could perform both actions simultaneously, and so the results cannot be attributed to pragmatic effects. One-action problems did not increase accuracy, but they yielded the fastest reading times for each premise and the fastest answers to the questions. The most plausible explanation is that one-action problems are the easiest to understand and to remember, because the same predicate occurs in all three assertions. This occurrence of the same predicate in each clause in a problem shortens reading time in comparison with one-actor problems, which merely have a referent common to each clause. Why didn't the one-action problems yield the most accurate inferences? The most likely explanation is that readers are confused by the similarity of all the clauses, and hence their models, as in the following example:

Karl is eating at the table or Sue is eating at the table but not both.

Karl is not eating at the table.

Is Sue eating at the table?

A similar distinction has been drawn in the literature on the fan effect. A stronger fan effect occurs when several people are described in one location than when one person is described in several locations. When the same information is stored in different models, there may be interference during retrieval (Radvansky, Spieler, & Zacks, 1993) and also during deductive inference.

EXPERIMENT 3

Co-reference is likely to have a striking effect on more complex disjunctions. Consider an exclusive disjunction of two conjunctions, such as the following:

Either Jane is kneeling by the fire and she is looking at the TV or otherwise Mark is standing at the window and he is peering into the garden.

Jane is kneeling by the fire.

Does it follow that she is looking at the TV?

The model theory predicts that naive reasoners will respond "yes." However, the inference is an illusion. To see why, and to understand the prediction, consider the general form of the disjunction:

Either P and Q or otherwise R and S.

The force of the exclusive disjunction is that one conjunction is true and the other is false, and so if *P and Q* is true then *R and S* is false, and vice versa. However, according to the principle of truth, naive individuals should construct just two mental models of such a disjunction:

```
P     Q
      R     S
```

These mental models predict that given the categorical premise *P*, reasoners should infer *Q*, and that they should also infer *not-R* (and *not-S*). Likewise, if the disjunction is combined with the categorical premise *not-P*, the mental models predict that reasoners should infer *not-Q* and that they should also infer *R* (and *S*). The mental models fail to represent what is false, but the fully explicit models of the disjunction do take falsity into account. Given the truth of one conjunction, the fully explicit models represent the different ways in which the other conjunction can be false. These models show that the disjunction is compatible with six possibilities:

```
P     Q     ¬ R     S
P     Q       R   ¬ S
P     Q     ¬ R   ¬ S
P   ¬ Q       R     S
¬ P   Q       R     S
¬ P ¬ Q       R     S
```

These fully explicit models allow us to correctly categorize the previous inferences, each of which depends on the same compound premise:

Either P and Q or otherwise R and S.

1. *P; therefore, Q* is an illusion (the fourth fully explicit model is a counterexample).
2. *P; therefore, not-R* is an illusion (the second model is a counterexample).
3. *Not-P; therefore, not-Q* is an illusion (the fifth model is a counterexample).
4. *Not-P; therefore, R* is valid (there are no counterexamples).

Co-reference and actions that are incompatible with one another modulate the fully explicit models and thereby make some of them impossible. We can illustrate these effects with the five sorts of disjunctions that we used in the present experiment as follows.

1. *One actor (incompatible actions from one conjunction to the other)*; for example,

Either Jane is kneeling by the fire and she is looking at the TV or otherwise she is standing at the window and she is peering into the garden.

If Jane is carrying out the two actions in the first conjunction, then she cannot be carrying out either of the two actions in the second conjunction. Hence, the assertion yields only two possibilities, which we abbreviate here in their fully explicit models (as is indicated at the bottom of this page). Hence, all four of the preceding inferences are valid for this sort of co-referential content.

2. *Two actors in each model (incompatible actions from one conjunction to the other)*; for example,

> Either Jane is kneeling by the fire and Mark is looking at the TV or otherwise Jane is standing at the window and Mark is peering into the garden.

Once again, if Jane is kneeling by the fire, then she cannot be standing by the window, and vice versa. Likewise, if Mark is looking at the TV, then he cannot be peering into the garden, and vice versa. Hence, there are again only two possibilities, and all four of the preceding inferences are valid for this sort of co-referential content too.

3. *Two actors in each model (compatible actions from one conjunction to the other)*; for example,

> Either Jane is kneeling by the fire and Mark is standing at the window or otherwise Jane is looking at the TV and Mark is peering into the garden.

In this case, it is possible for Jane to be both kneeling by the fire and looking at the TV; likewise, it is possible for Mark to be standing at the window and peering into the garden. Hence, the disjunction yields six possibilities (as is shown in the six fully explicit models above), and so the first three of the four preceding inferences are illusory for this sort of content.

4. *Two actors in separate models (compatible actions from one conjunction to the other)*; for example,

> Either Jane is kneeling by the fire and she is looking at the TV or otherwise Mark is standing at the window and he is peering into the garden.

As in the previous case, the disjunction yields six possibilities, and so the first three of the four preceding inferences are illusory for this sort of content.

5. *Four actors (compatible actions from one conjunction to the other)*; for example,

> Either Jane is kneeling by the fire and Sean is looking at the TV or otherwise Mark is standing at the window and Pat is peering into the garden.

The disjunction yields six possibilities, and so the first three of the four preceding inferences are illusory for this sort of content.

In each case, reasoners should rely on the two *mental* models. Hence, the theory predicts a uniform acceptance of the four preceding inferences. The disjunctions with incompatible actions serve as control problems, because the four inferences are all valid: They hold for the two possibilities that these disjunctions yield. The disjunctions with compatible actions should yield the same conclusions, but they are illusions for the first three of the four inferences; the conclusions do not hold in the six possibilities that these disjunctions yield.

Inferences should be made more readily in the *same* model condition (i.e., *P*, therefore *Q*) than in the *different* model condition (i.e., *P*, therefore *not-R*) because the former calls for consideration of only one of the mental models of the disjunction, whereas the latter calls for consideration of both mental models of the disjunction. However, the referential manipulation allows us to compare the effect of co-reference within the same model (co-reference in *P* and *Q*) with the effect of co-reference from one model to another (co-reference from *P* to *R*). Co-reference may enhance reasoning by yielding more concise initial models. However, co-reference may further enhance reasoning when the categorical premise and the conclusion in the question share co-referents. This manipulation allowed us to study whether drawing inferences about co-referents facilitates reasoning, independently of the amount of information held in mind.

Method

Participants. We tested individually 35 participants (25 paid members of the public recruited through national newspaper advertisements and 10 postgraduate volunteers from the University of Dublin, Trinity College). The participants were 14 men and 21 women ranging in age from 18 to 78 years, with a mean age of 36 years.

Design and Materials. Each participant acted as his or her own control. Each trial was based on a disjunction (*Either P and Q or otherwise R and S*), and its first clause, *P*, was either asserted or denied. Same-model problems occurred on half the trials: The participants were asked if *Q* follows. Different-model problems occurred on the other half of the trials: The participants were asked if *R* follows. As we described earlier, there were five sorts of reference in the disjunctions. Two sorts had actions that were incompatible from one model to another: one-actor and two-actor disjunctions, which yield only two possibilities and, hence, valid inferences for the four sorts of inference. Three sorts had actions that were compatible from one model to another: disjunctions with two actors in each model, disjunctions with two actors in separate models, and disjunctions with four actors, which yield six possibilities and, hence, illusions in three of the four inferences. Each participant carried out the four forms of inference with each of the five sorts of disjunction in a different random order, and the 20 inferences had different lexical contents.

We constructed 20 sets of contents concerning four different common actions, each containing the same number of words. In the experiment as a whole, these materials were rotated so that with the appropriate referential adjustments they were presented equally often with each of the five sorts of referential problems. We selected

Jane:

Kneeling by fire	Looking at TV	¬Standing at window	¬Peering into garden
¬Kneeling by fire	¬Looking at TV	Standing at window	Peering into garden

a sample of 16 premises, 8 of which referred to compatible actions and 8 to incompatible actions. This categorization was corroborated empirically. The 24 participants in the norming study for Experiment 1 rated conjunctions of the compatible actions as possible on 98% of the trials, and they rated conjunctions of the incompatible actions as possible on only 5% of the trials (all the participants conformed to this difference in their ratings; binomial test, $p = .5^{24}$).

Procedure. The problems were presented on a computer screen using the E-Prime package. The participants read on-screen instructions, which were the same as those in Experiment 2. They were not told that their responses would be timed. They then completed 5 practice problems followed by the 20 experimental problems. For each problem, the premises and questions were presented one by one on the screen. After reading each premise, the participants pressed a key to access the next premise and, finally, the question. The premises remained on the screen until the participants responded by pressing one of three keys: "yes," "no," or "cannot tell" (corresponding to the T, O, and U keys, respectively). The program recorded the reading times for each of the premises and the time taken to respond to the question.

Results and Discussion

The percentages of the predicted inferences and their mean latencies to the main *forms* of problems are presented in Table 5. Because there were so many predicted, though erroneous, responses to the illusory problems, the table presents the latencies of the predicted responses rather than those of the correct responses. As the table shows, most inferences fit the predictions of the model theory: 28 participants made more predicted than unpredicted inferences, 4 participants made fewer predicted than unpredicted inferences, and the remaining 3 participants had ties (binomial test, $p < .0005$). As a corollary, the participants performed well with the control problems, which were based on incompatible actions allowing only two possibilities (78% correct). In contrast, they performed badly with the illusory problems, which were based on compatible actions allowing six possibilities (only 10% correct). Thirty-four of the 35 participants were more accurate on the control problems than on the illusory problems (binomial test, $p < 1$ in 900 million). There was no reliable difference in the percentages of predicted inferences to the control problems and to the illusory problems (Wilcoxon test, $z = 1.67, p < .1$, two-tailed). Hence, the participants are likely to have constructed the two predicted *mental* models for both sorts of problems regardless of the number of possibilities that the disjunctions yield.

As Table 5 shows, the participants responded faster in the same-model conditions ($M = 7.1$ sec) than in the different-model conditions ($M = 9.8$ sec; Wilcoxon test, $z = 4.86$, $p < .0005$). Hence, the speed of an inference depends on whether or not reasoners can draw a conclusion from the same model referred to by the categorical premise. The participants also responded faster when the categorical premise was affirmative (i.e., *P*) and, hence, matched a mental model of the disjunction ($M = 8.2$ sec) than when it did not (i.e., *not-P*; $M = 8.8$ sec; Wilcoxon test, $z = 2.54$, $p < .006$). The interaction was also significant (Wilcoxon test, $z = 2.1, p < .04$, two-tailed: The increased latencies in the latter case were larger in the same-model condition than in the different-model condition. Hence, reasoners were faster when the predicted answer was explicitly represented in their initial model of the premise—that is, when the answer was affirmative.

The percentages of predicted responses, reading times, and times for the predicted responses for the five referentially distinct sorts of disjunction are presented in Table 6. As in the previous experiments, co-reference made the process of inference easier. The participants made more predicted than unpredicted responses over the three sorts of problems: one-actor problems (84%), two-actor problems (80%), and four-actor problems (75%). The trend was only marginal (Page's $L = 430.5$, $z = 1.26, p < .11$), but the difference between one-actor and four-actor problems was reliable (Wilcoxon test, $z = 1.80, p < .04$). No other referential effects on percentages of predicted responses were reliable.

The times taken to read the five different sorts of disjunctive premise did not differ reliably (Friedman test, $\chi^2 = 4.28, p = .37$). However, there was a marginal difference in reading times of the categorical premise (Friedman test, $\chi^2 = 8.57, p < .08$). These reading times were shorter when the categorical premises occurred with disjunctions that had co-reference within each model ($M = 5.0$ sec) than when they occurred with disjunctions with different referents within each model ($M = 5.5$ sec). The response times across the five different sorts of disjunctive problem also differed reliably both for all responses [Friedman test, $\chi^2(1) = 20.08, p = .0005$] and for the responses predicted by the mental model theory [Friedman test, $\chi^2(1) = 10.8, p < .03$]. It is hard to know what the participants were doing when they made unpredicted responses. However, when they made predicted responses, they were faster when the disjunctive premise referred to one actor ($M = 6.8$ sec) than when it referred to more than one actor ($M = 8.9$ sec;

Table 5
Percentages of Predicted Inferences and Their Latencies (RTs, in Seconds) for the Control and Illusory Problems in Experiment 3

	Categorical Premise							
	Same Model		Different Model		Same Model		Different Model	
Predicted Inference	P ∴ Q	RT	P ∴ Not-R	RT	Not-P ∴ Not-Q	RT	Not-P ∴ R	RT
Control problems (two possibilities)	84	6.7	87	12.2	77	7.5	83	10.4
Illusory problems (six possibilities)	87	5.8	69	8.3	75	8.2	78*	7.9
Overall means		6.4		10.0		7.9		9.7

*Only this inference is valid for the illusory problems.

Table 6
Percentages of Predicted Responses, Reading Times, and Times for the Predicted Responses for the Five Sorts of Disjunction in Experiment 3

Semantic Content	% Predicted Responses	Time (Sec)			
		Reading of Disjunctive Premise	Reading of Categorical Premise	Response	
				Same Model	Different Model
Control problems (two possibilities):					
1. One actor	84	10.4	5.2	6.5	7.3
2. Two actors in each model	82	11.8	5.6	7.8	9.8
Illusory problems (six possibilities):					
3. Two actors in each model	81	10.8	5.5	7.4	9.8
4. Two actors in separate models	75	9.9	4.7	7.1	11.9
5. Four actors	75	11.3	5.1	6.7	10.8

Wilcoxon test, $z = 3.57, p < .001$). When the disjunction referred to one actor, the difference between the same-model and the different-model conditions was smaller than when it referred to more than one actor (Wilcoxon test, $z = 2.21, p < .03$). Hence, problems that require reasoners to consider both mental models are easier when the same individual occurs in both of them. No other interactions were reliable.

GENERAL DISCUSSION

Co-reference can improve reasoning. In Experiment 1, we examined biconditionals and logically equivalent exclusive disjunctions, and in both cases the participants were more accurate in reasoning from assertions with one co-referential actor, such as

> If and only if Rita is not looking into the wardrobe then Rita is sweeping under the table

than from assertions about two actors, such as

> If and only if Alan is not looking under the bed then Cathy is washing at the sink.

The model theory predicted the advantage of one-actor assertions. Background knowledge reduces the plausibility that both actions occur together, and the mental models represent only three referents, whereas the mental models of two-actor assertions represent four referents.

In Experiment 2, it was confirmed that one-actor disjunctions yield more accurate inferences than two-actor disjunctions. Because the actor could perform both actions at the same time, the difference cannot be attributed to pragmatics. The result accordingly corroborates the model theory's claim that co-reference yields simpler models. The experiment also showed that individuals read the premises and respond fastest to disjunctions in which two actors carry out the same action. However, one-action disjunctions yield less accurate inferences than one-actor disjunctions. In our view, the explanation for this is that the speed of response reflects the occurrence of the same predicate in all three assertions—the two premises and the conclusion—but this similarity also confuses reasoners (cf. Zwaan & Radvansky, 1998).

In Experiment 3, we examined inferential problems based on disjunctions of conjunctions, such as

> Either Jane is kneeling by the fire and Sean is looking at the TV or otherwise Mark is standing at the window and Pat is peering into the garden.
>
> Jane is kneeling by the fire.
>
> Does it follow that Sean is looking at the TV?

The majority of the participants responded "yes," but the inference is an illusion. It is entirely possible that Jane is kneeling at the fire and that Sean is *not* looking at the TV, provided that the second conjunction is true: Mark is standing at the window and Pat is peering into the garden. The error is predictable from the model theory's principle of truth: Individuals think about the truth of the first conjunction without considering the falsity of the second conjunction, and vice versa. Hence, they envisage just two possibilities as compatible with the disjunction shown at the bottom of this page. Such inferences have the force of cognitive illusions and are resistant to various sorts of potential antidotes (see, e.g., Johnson-Laird & Savary, 1999).

Co-reference had two effects on the inferences in Experiment 3. First, it made the process of inference easier. The proportions of predicted responses increased from four-actor problems to two-actor problems, and further from two-actor problems to one-actor problems. The times to make the predicted responses also declined according to the same trend. Second, co-reference and the content of clauses can modulate the interpretation of sentential connectives (see Johnson-Laird & Byrne, 2002). When the disjunction concerned only one actor, as in

> Either Jane is kneeling by the fire and she is looking at the TV or otherwise she is standing at the window and she is peering into the garden,

Jane: kneeling by fire Sean: looking at TV

Mark: standing at window Pat: peering into garden

the content of the clauses is inconsistent with certain possibilities. It is impossible, for instance, for Jane to be both kneeling by the fire and standing at the window. Hence, a reasonable interpretation of the sentence is that it is compatible with only two possibilities. In this case, the inferences are no longer illusory. For example, given the further premise:

Jane is kneeling by the fire

the following conclusion:

She is looking at the fire

is valid.

In sum, the results show that co-reference can aid reasoning. Following Bouquet and Warglien (1999), our results have extended the effects to a wider variety of sorts of assertion and sorts of inference, and they have shown that they are robust and statistically significant. When a compound premise had only a single actor, reasoners produced more of the predicted responses. Similarly, co-reference can speed up inferences. Disjunctions with only one actor yielded faster responses than disjunctions with two or more actors. In Experiment 3, the effect was strongest in the different-model condition—that is, when the reasoners had to consider an alternative model to the one representing the categorical premise. They therefore had to think about two models, which places a bigger load on working memory. This difficulty in thinking about two models is reduced when the categorical premise and the question refer to the same individuals.

The occurrence of the illusory inferences and the effects of co-reference are embarrassing to theories of reasoning based on formal rules of inference (see Braine & O'Brien, 1998; Rips, 1994). In order to account for the sentential inferences in Experiments 1 and 2, these theories invoke formal rules, such as

P or Q

Not-P

Therefore, Q.

Such rules are blind to whether *P* and *Q* have a referent in common, and so these theories have no immediate way of explaining the effects of co-reference. Given premises of the form

Either P and Q or otherwise R and S

P,

the vast majority of participants in Experiment 3 drew the conclusion

Q.

This conclusion is not formally valid, and so formal rule theories have no immediate way of explaining why the participants drew this conclusion; the theories rely on valid rules of inference. It may be feasible to modify the theories to meet these two challenges, but the required modifications are not obvious.

The theory of mental models predicts that co-reference should improve reasoning by yielding simpler mental models. The difficulty of reasoning depends on the number of models required to solve a problem, but it also depends on the complexity of those models. The theory predicts that illusory inferences should occur when falsity matters. It also predicts that co-reference, semantics, and general knowledge can modulate the interpretation of sentential connectives. Such modulations eliminate models of possibilities that would normally be neglected, and thereby transform an illusory inference into a valid one.

REFERENCES

Barres, P. E., & Johnson-Laird, P. N. (2003). On imagining what is true (and what is false). *Thinking & Reasoning*, **9**, 1-42.

Bouquet, P., & Warglien, M. (1999). Mental models and local models semantics: The problem of information integration. In *Proceedings of the European Conference on Cognitive Science* (pp. 169-178). Siena, Italy.

Braine, M. D. S., & O'Brien, D. P. (1998). *Mental logic*. Mahwah, NJ: Erlbaum.

Garnham, A., Oakhill, J., & Johnson-Laird, P. N. (1982). Referential continuity and the coherence of discourse. *Cognition*, **11**, 29-46.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Cambridge University Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove, U.K.: Erlbaum.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics and inferences. *Psychological Review*, **109**, 646-678.

Johnson-Laird, P. N., & Savary, F. (1996). Illusory inference about probabilities. *Psychologica*, **93**, 69-90.

Johnson-Laird, P. N., & Savary, F. (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition*, **71**, 191-229.

Radvansky, G. A., Spieler, D. H., & Zacks, R. T. (1993). Mental model organization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 95-114.

Rips, L. J. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.

Schaeken, W., Johnson-Laird, P. N., Byrne, R. M. J., & d'Ydewalle, G. (1995). A comparison of conditional and disjunctive inferences: A case study of the mental model theory of reasoning. *Psychologica Belgica*, **35**, 57-70.

Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral & Brain Sciences*, **16**, 417-494.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, **123**, 162-185.

(*Continued on next page*)

APPENDIX
Problems Based on Exclusive Disjunctions in Experiment 1

One Actor

Rita is looking into the wardrobe or Rita is sweeping under the table but not both.
Rita is looking into the wardrobe.

Mary is reading beside the lamp or Mary is swimming in the pool but not both.
Mary is not reading beside the lamp.

Barry is working at the bench or Barry is kneeling on the rug but not both.
Barry is kneeling on the rug.

Sarah is sitting in the armchair or Sarah is opening the front door but not both.
Sarah is not opening the front door.

Two Actors

Brian is standing by the fireplace or Joanne is looking in the mirror but not both.
Brian is standing by the fireplace.

Rachel is climbing up the stairs or David is cooking at the stove but not both.
Rachel is not climbing up the stairs.

Alan is looking under the bed or Cathy is washing at the sink but not both.
Cathy is washing at the sink.

Karen is sitting in front of the TV or Louise is sleeping in the bed but not both.
Louise is not sleeping in the bed.

Graham is standing on the scales or Carol is standing on the scales but not both.
Graham is standing on the scales.

Michael is writing at the desk or Alex is writing at the desk but not both.
Michael is not writing at the desk.

Eric is sitting in the bath or Daniel is sitting in the bath but not both.
Daniel is sitting in the bath.

Martin is standing on the stool or Susan is standing on the stool but not both.
Susan is not standing on the stool.

Ruth is eating at the table or Alice is eating at the table but not both.
Ruth is eating at the table.

Linda is looking out the window or Richard is looking out the window but not both.
Linda is not looking out the window.

Peter is sitting on the sofa or Andrew is sitting on the sofa but not both.
Andrew is sitting on the sofa.

Mark is leaning on the counter or Lisa is leaning on the counter but not both.
Lisa is not leaning on the counter.

(Manuscript received December 18, 2002;
revision accepted for publication August 19, 2003.)