The Semantic Modulation of Deductive Premises

Clare R. Walsh (cwalsh@tcd.ie)
Psychology Department, University of Dublin, Trinity College,
Dublin 2, Ireland

P.N. Johnson-Laird (phil@princeton.edu)
Department of Psychology, Green Hall, Princeton University,
Princeton, NJ 08544, USA

Abstract

Two experiments examined how the mental models of premises influence deductive reasoning. Experiment 1 showed that individuals draw different conclusions from the same information depending on whether it is expressed in conditional assertions or disjunctions. It also showed that co-reference within the premises can speed up more difficult inferences. Experiment 2 corroborated these results and also showed that the failure to represent what is false can lead people to draw illusory inferences, i.e., systematic but compelling fallacies.

Introduction

The ability to reason deductively is central to human intelligence. Our goal is to examine how the nature of mental representations can influence this ability. Several factors should influence the process. First, the verbal formulation of premises should lead to different representations of the same underlying information. In turn, these representations should lead reasoners to draw different conclusions. Second, the semantic content of premises should influence the process of deductive reasoning. It should eliminate certain possibilities, and, in the case of co-reference from one clause to another, it may yield more concise representations. In this paper, we examine how these different factors impact on deductive reasoning.

Mental Models and Deductive Reasoning

The theory of mental models postulates that individuals who have no training in logic represent the meaning of assertions in mental models (Johnson-Laird and Byrne, 1991). Each mental model represents a possibility. But, the limitations of working memory force individuals to abide by the *principle of truth:* mental models represent only true possibilities and within them the constituent propositions in premises only when they are true. For example, an exclusive disjunction such as:

Either Mary is in Brussels or Gino is in Rome, but not both.

elicits two mental models of the alternative possibilities:

Mary-Brussels

Gino-Rome

where "Mary-Brussels" denotes a model of Mary in Brussels, and "Gino-Rome" denotes a model of Gino in Rome. Mental models of possibilities do not represent the falsity of the clauses in the disjunction, e.g., that Gino is not in Rome in the first possibility. One consequence is that mental models should give rise to illusory inferences, i.e., fallacies that most individuals make. Recent studies have corroborated their occurrence (e.g. Johnson-Laird and Savary, 1999). In one experiment, participants were given the following problem:

If there is a king in the hand then there is an ace in the hand, or else if there is not a king in the hand then there is an ace in the hand.
There is an king in the hand.
What follows?

All the participants drew the conclusion that there is an ace in the hand. Mental models yield this conclusion, but it is wrong. In fact, the sentential connective or else implies that one of the conditionals could be false. But, if the first conditional is false, then there isn't an ace in the hand even though there is a king in the hand.

The meanings of clauses, their co-referential relations, and background knowledge, can all *modulate* the basic meanings of sentential connectives (Johnson-Laird and Byrne, 2001). One way modulation can occur is when background knowledge prevents the construction of a model. Consider, for example, the conditional:

If she played a game, then she didn't play soccer.

The basic interpretation of a conditional allows the possibility that she doesn't play a game but that she plays soccer. This possibility, however, is eliminated by the knowledge that soccer *is* a game.

A second example of modulation, in our view, is Bouquet and Warglien's (1999) discovery that

reasoning from disjunctions was more accurate when the clauses were co-referential, e.g.:

Either Gino is in Brussels or Gino is in Rome, but not both. Gino is not in Rome. What follows?

Premises of this sort yielded a greater proportion of valid conclusions, e.g.: Gino is in Brussels, than premises that did not refer to the same individual in both clauses. Co-reference may enable people to construct representations that are more concise. The aim of Experiment 1 was to examine this possibility.

Experiment 1: Co-reference in reasoning

Our conjecture is that co-reference may allow reasoners to build a more concise representation of the premises, thereby reducing the load on working memory. We predicted that such representations should facilitate performance in reasoning tasks particularly if the demands on working memory are high. In addition, the form of the premises - whether they are based on a disjunction or a conditional - should make different information available. Hence, each sort of premise should make some inferences easier and some more difficult to draw.

Each participant carried out 32 inferences in a different random order. The inferences concerned people carrying out various actions. For half the inferences, the first premise was an exclusive disjunction, e.g.:

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

which should elicit the mental models:

Rachel-climbing

David-cooking

For half the inferences, the same information was expressed as a biconditional, e.g.:

If and only if Rachel is not climbing up the stairs then David is cooking at the stove.

which should elicit the mental models:

¬ Rachel-climbing David-cooking

. . .

where "¬" denotes negation, and the ellipsis denotes an implicit model, which acts as a "place holder" for the possibilities in which the antecedent is false. If necessary, it can be fleshed out explicitly.

The second premise was a categorical assertion or categorical denial of either the first or second proposition in the preceding premise. There were accordingly 8 forms of inference.

In order to manipulate co-reference, there were four types of semantic content:

1. Two persons do two different actions (two-actions).
2. Two persons do one action, which they cannot perform simultaneously (exclusive-action, e.g. "sit on the stool").
3. Two persons do one action, which they can perform simultaneously (inclusive-action, e.g., "sit on the sofa").
4. One person does two different actions, which cannot be performed simultaneously (one-person).

The eight forms of inferences with the four sorts of content yielded the 32 different inferences.

The problems were presented on a computer. The participants drew their own conclusion about what followed from the premises. They responded by typing their answer, and their latencies were measured from the presentation of the premises to the first key press. They were not told that their responses were being timed. We tested 30 undergraduates from Princeton University in return for course credit.

Results and Discussion

Table 1 presents the percentages of correct responses to the eight forms of inference (collapsing over their contents). As the Table shows, the participants were more accurate when the categorical premise matched an event that was represented explicitly in the models of the first premise. When the categorical premise was negative and concerned the first event in the preceding premise, Not-A, the participants were more accurate in reasoning from the biconditional (91%) than the disjunction (68%). But, when the categorical premise was affirmative, A, they were more accurate in reasoning from the disjunction (94%) than from the biconditional (78%; McNemar tests, chi-squared = 17.36 $p < .0001$; chi-squared = 15.04, $p < .0001$ respectively). Participants performed equally well on both descriptions when the categorical premise affirmed the second proposition, B, (94% for both, McNemar Test, chi-squared = 0.07 $p = 1$). We predicted these results in terms of mental models, but they might reflect the surface matching of clauses in the premises. One result, however, is more readily explained in terms of models. When the categorical premise negated the second proposition, not-B, participants were more accurate in reasoning from the disjunction (68%) than from the biconditional (46%; McNemar Test, chi-squared = 13.8, $p < .0002$). Not-B mismatches the clauses in both sorts of premises. It is also not represented explicitly in the initial models of either premise. But, reasoners may find it easier to flesh out the disjunction, which already contains two mental models, than to flesh out the conditional which is represented by just one explicit mental model.

Table 1
The percentages of correct responses to the eight forms of inference in Experiment 1

	First Premise	
Categorical Premise	*Iff not-A then B*	*A or else B*
A	78	94
Not-A	91	68
B	94	94
Not-B	46	68

The co-referential manipulation had no effect on accuracy. Yet, it did affect the latencies of correct responses. The principal results were that responses to exclusive-action problems (10.93 secs) and to inclusive-action (10.18 secs) were faster than those to two-action problems (11.58 secs; Wilcoxon test $z = 2.33, 2.81$, $p < .01$, $p < .003$, respectively, excluding the results of two outliers). In other words, co-reference can help the process of reasoning. In particular, problems in which two persons carry out one action, whether or not they can perform it at the same time, elicited faster responses than problems in which two persons carry out two different actions. Reference to a common action may yield more parsimonious models of the premises, and it may help reasoners to avoid confusion about which action a particular individual was carrying out. Such confusions are more likely in the case of a disjunction, which, unlike a biconditional, demands that reasoners model two explicit possibilities. The shared referent in a disjunction is common to two alternative models whereas in a biconditional it occurs within one model. That, perhaps, is why the referential effects were stronger for disjunctions (exclusive-action 9.0 secs vs two-action problems 10.7 secs, inclusive-action 8.7 secs vs two-action problems 10.6 secs, Wilcoxon test $z = 2.76, 2.44$, $p < .003, .01$ respectively), than biconditionals (exclusive-action 11.8 secs vs two-action problems 12.3 secs, inclusive-action 11.7 secs vs two-action problems 12.0 secs, Wilcoxon test $z = 0.68, 1.47$, $p < .25, .07$ respectively). We followed up these phenomena in a second experiment.

Experiment 2: Illusory inferences and co-reference

The aims of the experiment were twofold. The first aim was to examine what inferences people make from an exclusive disjunction of the form:

Either P and Q or otherwise R and S.

The disjunction was paired with a categorical premise, either asserting or denying its first proposition, P. The participants had to evaluate the validity of a one-clause conclusion, either Q or R. There were accordingly four forms of inference.

The model theory predicts that the failure to make certain information explicit in the models of the disjunction should lead people to make invalid inferences. According to the principle of truth, reasoners should construct two mental models of such a premise:

P Q
R S

It follows that given the categorical premise P, reasoners should infer that Q and not-R follow. Similarly, given the categorical premise not-P they should infer that not-Q follows. These inferences, however, are illusions. When falsity is taken into account, the disjunctive premise is consistent with six different possibilities, which we present here in *fully explicit* models:

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 models show that the three previous inferences are invalid. Given the premise, P, for instance, Q and not-Q are both possible, and likewise R and not-R are both possible. In contrast, given the categorical premise, not-P, participants should correctly infer that R follows: this conclusion follows from the mental models above, but it also follows from the fully explicit models.

Our second aim was to examine the effects of co-reference on these inferences. Experiment 1 showed that co-reference reduced response times, at least for certain inferences. The present experiment followed up this effect and the semantic modulation of the premises.

Table 2 presents the five sorts of semantic contents, which manipulated the number of shared referents (i.e., people carrying out actions) contained in the first premise, and whether the co-referential relations occurred within or between models. We predicted that co-reference would again facilitate performance and that this facilitation would be greater as the number of shared referents increased. We also predicted that facilitation would be greater for problems requiring a greater working memory load.

Table 2: The five sorts of semantic content in Experiment 2, and the number of fully explicit models compatible with each content.

1. *Four persons act: six models*

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.

2. Two persons, one per clause: six models

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.

3. Two persons do inclusive actions: six models

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.

4. Two persons do exclusive actions: two models

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.

5. One person: two models

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.

The first three sorts of content in Table 2 are consistent with all six of the fully explicit models above. However, in the other two cases, the exclusive actions rule out the models in which P and R occur together, and Q and S occur together. The premise is therefore consistent with just two fully explicit models:

P	Q	¬ R	¬ S
¬ P	¬ Q	R	S

These models yield the same conclusions as the mental models described above, but these conclusions are no longer illusions, but correct.

We tested individually 35 participants (25 paid members of the public and 10 postgraduate volunteers from the University of Dublin, Trinity College). They acted as their own controls and carried out the 20 inferences in different random orders. We constructed 20 sets of materials, each of which contained the same number of words. The materials were rotated so that they were presented equally often with each of the five sorts of semantic content.

The problems were presented on a computer screen. For each problem, the premises and questions were presented one-by-one on the screen. Participants responded to the question by pressing one of three keys: yes, no or cannot tell. The program recorded separately the time that it took participants to read each of the premises and to answer the question. The participants were not told that their responses were being timed.

Results and Discussion

Table 3 presents the percentages of the "Yes" and "No" responses to the main sorts of problems. As the table shows, the participants succumbed to the illusory (six model) problems (10% correct), but performed well the (six-model) control problem (78% correct). Of the 35 participants, 34 were less accurate on the illusions than on the control problems (Sign Test, $p < 1$ in 900 million). The pattern of responses was comparable for the two model problems, and so we suspect that the participants constructed just two models for the six-model and the two-model problems. A similar pattern of results occurred in the participants' responses to the first problem that they encountered. They were more accurate with the two-model problems (81% correct) than with the six-model problems (31% correct). This result suggests that the illusions occurred spontaneously and not as a result of the development of a mental set.

Table 3: The percentages of the "Yes" and "No" responses in Experiment 2. The balance of the responses (around 10% per problem) were "can't tell". The predicted responses are shown in bold, and underlined where they are illusory.

	Six models		Two models	
1. P & Q, or R & S	Yes	No	Yes	No
P				
∴ Q	**87**	4	**84**	6
2. P & Q, or R & S				
P				
∴ R	22	**69**	7	**87**
3. P & Q, or R & S				
Not-P				
∴ Q	13	**75**	12	**77**
4. P & Q, or R & S				
Not-P				
∴ R	**78**	15	**83**	10

There was no reliable difference between the five semantic conditions in response accuracy or in the time that it took to read the first premise. But, a significant difference in response times occurred across conditions when all responses were considered (Friedman Test, chi-squared = 20.08, df = 1, $p = .0005$), and when we included only the responses predicted by the mental model theory (Friedman Test, chi-squared = 10.66, df = 1, $p = .03$). All further analyses are based only on the responses predicted by the mental model theory, because it is difficult to know what the participants were doing when they got the answer right to the illusory problems and wrong to the control problems.

There were two principal results:

1. Responses were faster when the first premise referred to fewer individuals (1-person condition, mean 6.8 secs, 2-person conditions, mean = 8.6 secs, and 4-person condition, mean = 8.3 secs; Kruskall Wallis Test, chi-squared = 10.42, df = 2, $p < .003$).

2. The difference was significant in the different-model condition (1-person condition, mean 7.4 secs, 2-person conditions, mean = 9.9 secs, and 4-person condition, mean = 10.5 secs; Kruskall Wallis Test, chi-squared = 12.2, df = 2, $p < .002$) but not in the same-model condition (1-person condition, mean 6.3 secs, 2-person conditions, mean = 7.4 secs, and 4-person condition, mean = 6.6 secs; Kruskall Wallis Test, chi-squared = 1.89, df = 2, $p < .20$).

The results corroborated the occurrence of illusory inferences, and reasoners seem likely to construct just two models of disjunctions of the form:

Either P and Q or otherwise R and S.

They overlook the different ways in which the conjuncts could be false in the case of the six model problems, i.e., those with a content that does not eliminate any of the possibilities. The latencies of the responses bear out our conjecture that inferences are easier when the premises concern fewer individuals. Reasoners can construct more concise models in this case and are less open to confusion about who did what. Faster responses occurred both when the co-referential relation was within one model and when it was between items in different models. However, response times were faster only when reasoners had to consider an alternative model to the one referred to in the categorical premise, probably because that condition places an extra load on working memory.

General Discussion

The mental model theory predicts that when people represent a premise, they do so in accordance with the principle of truth. They construct a representation that makes only some information explicit. Experiment 1 corroborated the principle. It showed that reasoners draw different conclusions from given information, depending on whether it is expressed as a biconditional or an exclusive disjunction. Likewise, Experiment 2 showed that the failure to represent falsity can lead reasoners to make illusory inferences.

A second factor influences mental models: the occurrence of co-reference within the premises. Experiment 1 showed that co-reference within an exclusive disjunction speeded up the process of inference. Experiment 2 showed the same effect, but only when reasoners had to consider an alternative model to the one referred to in the categorical premise. The inferential task in such cases places a bigger load on working memory, and co-reference evidently ameliorates matters. The same factor may explain why there was no facilitation for conditionals in our first experiment. Reasoners can construct a more concise representation of premises containing co-referents. This parsimony reduces the load on working memory and the latencies of more difficult inferences.

Acknowledgements

Clare Walsh is supported by an Enterprise Ireland Ph.D. Fellowship and a Government of Ireland Scholarship for the Humanities and Social Sciences, and a Dublin University Postgraduate Award. We thank Ruth Byrne, Eugenia Goldvarg, Uri Hasson, Markus Knauff, Yingrui Yang, and Lauren Ziskind, for their helpful comments on this research.

References

Bouquet, P. & Warglien, M. (1999). Mental Models and Local Models Semantics: the Problem of Information Integration. *Proceedings of the European Conference on Cognitive Science* (ECCS'99), pp.169--178, Siena, Italy.

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

Johnson-Laird, P.N. & Byrne, R.M.J. (2001). Conditionals: a theory of meaning, pragmatics and inference. *Submitted.*

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