The Figural Effect and a Graphical Algorithm for Syllogistic Reasoning*

Peter Yule
Centre for Cognitive Science
Edinburgh University
U.K.
pgy%cogsci.ed.ac.uk@nsfnet-relay.ac.uk

Keith Stenning
Human Communication Research Centre
Edinburgh University
U.K.
keith%cogsci.ed.ac.uk@nsfnet-relay.ac.uk

Abstract

Theories of syllogistic reasoning based on Euler Circles have foundered on a combinatorial explosion caused by an inappropriate interpretation of the diagrams. A new interpretation is proposed, allowing single diagrams to abstract over multiple logical models of premisses, permitting solution by a simple rule, which involves the identification of individuals whose existence is entailed by the premisses. This solution method suggests a performance model, which predicts some of the phenomena of the Figural Effect, a tendency for subjects to prefer conclusions in which the terms preserve their grammatical status from the premisses (Johnson-Laird & Steedman 1978). 21 students were asked to identify the necessary individuals for each of the 64 pairs of premisses. The order in which the three terms specifying the individuals were produced was shown to be as predicted by the performance model, but contrary to the presumed predictions of Mental Models theory.

Introduction

Syllogisms are arguments from two premisses to a conclusion. Both premisses and conclusion are statements of one of four types: "All of the A are B" (A), "Some of the A are B" (I), "None of the A are B" (E) and "Some of the A are not B" (O). Each statement in the premisses contains two terms: one term, the middle term (b), occurs in both premisses, while the other two (a and c) are known as the end terms. The arrangement of the end and middle terms in each of the premisses gives rise to a four-way classification, known as the figure of the syllogism (see Table 1). It should be noted that our terminology in this matter follows the usage of Johnson-Laird (1983), and that like Johnson-Laird, we take the term "syllogism" to refer to any of the 64 pairs of premisses, rather than a pair of premisses plus a conclusion.

Psychological theories of syllogistic reasoning attempt to account for subjects' errors and response biases in solving syllogisms. Although early theories such as the Atmosphere Hypothesis (Woodworth and Sells 1935) account for some common errors, they cannot account for correct performance in many cases. Theories which do explain correct performance are often based on the logical model theory of the syllogism. Most serious theories of this type are based on the method of Euler Circles, which uses circles to represent sets of entities.

Unfortunately the term "method" is perhaps a misnomer for the use of Euler Circles, since the exact method of their employment is at best implicit in logic textbooks. Psychological theories typically have used each diagram to represent a single logical model of the premisses, resulting in a many-many mapping between statement types and diagrams (see Fig. 1). Two diagrams, one for each premiss, are integrated to form a ***registration diagram*** by superimposing the middle term circles from each. To solve a syllogism this way, all possible combinations of all possible pairs of diagrams have to be considered. A valid conclusion is then one which is modelled by all of the possible registration diagrams for a problem. The theories of Erickson (1974) and Sternberg (e.g. Guyote and Sternberg 1981) both make this assumption. Unfortunately the number of registration diagrams can be inordinately large, and the methods employed by these theories to cut down the search space make them incomplete and

Premiss	*Figure* 1	2	3	4
1	ab	ba	ab	ba
2	bc	cb	cb	bc

Table 1: The four figures of the syllogism.

*The support of the Economic and Social Research Council UK (ESRC) is gratefully acknowledged, both for their funding of the Human Communication Research Centre (HCRC), and the support given to Peter Yule in the form of an ESRC Research Studentship Award.

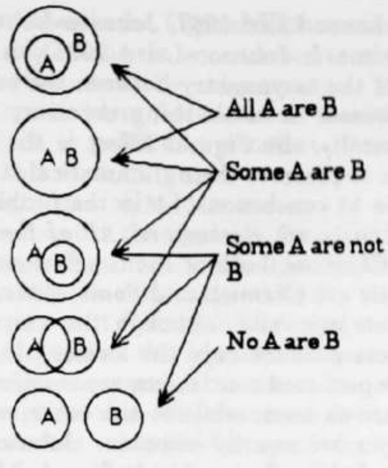

Fig. 1: The five Gergonne relations between a pair of circles, and how they model each of the four statement types.

consequently unable to account for correct performance (Johnson-Laird 1983).

Johnson-Laird's rejection of all methods based on Euler circles rests on a tacit assumption that Euler circles must be 'primitively' interpreted, so that one diagram always stands for one logical model. His proposed alternative is the theory of Mental Models, in which individual mental representations stand for more than one logical model, by devices such as the representation of optional elements (Johnson-Laird 1983), or a convention that makes it explicit when a set has been exhaustively represented, constraining the ways in which a skeletal model can be "fleshed out" (Byrne and Johnson-Laird 1991). Either way, each mental model abstracts over one or more logical models, so that an explosion of mental representations is avoided.

But Euler circles can also be interpreted in such a way that single diagrams abstract over multiple logical models. Each statement type has a *maximal model*, which contains all individuals which are compatible with the statement, and a *minimal model*, which contains only individuals whose existence is entailed by it. Representing the minimal model as a shaded region within the diagram which represents the maximal model, we obtain just one *characteristic diagram* per statement (see Fig. 2).

Each region within the diagram represents a different type of individual. A simple notation for individuals is the *type descriptor*. This is just a feature structure. Features are constructed from the terms appearing in the premisses, prefixed by "+" or "−", which indicate whether the individual concerned is or is not a member of the set denoted by the term. For example "+sentient-creature −Martian", describes a sentient creature which is not Martian.

The set of individuals whose existence is entailed by the premisses (*necessary individuals*) can be de-

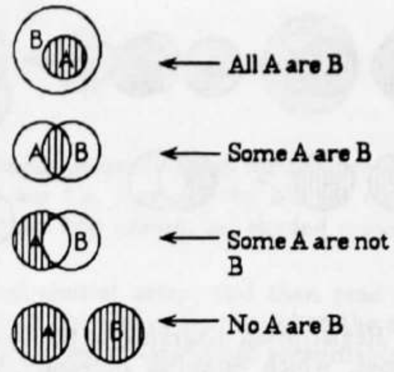

Fig. 2: Characteristic diagrams for each statement type, with shaded regions representing minimal models.

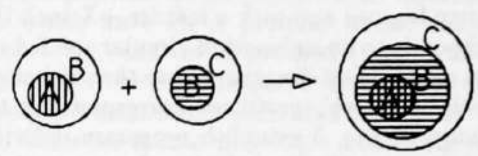

Fig. 3: Registration diagram for the problem "All of the a's are b's, All of the b's are c's". The central, unbisected circle corresponds to the individual +a+b+c.

termined by forming a registration diagram from the two characteristic diagrams, overlapping the circles representing the end terms if this is consistent with the premisses (this is equivalent to forming a maximal model of the premiss pair). The necessary individuals correspond to shaded regions from the characteristic diagrams which are not bisected during the formation of the registration diagram. Fig. 3 shows an example registration diagram, for a problem which establishes a necessary individual, and Fig. 4 shows the registration diagram for a problem which does not establish any necessary individuals, and so lacks valid conclusions. In all there are 21 distinct registration diagram types, the full set of which can be found in Stenning (1992).

Necessary individuals can form the basis for quantified conclusions. Particular (i.e. existential) conclusions can be drawn immediately by dropping the middle term from the type descriptor and picking a + term as the subject of the conclusion (e.g.

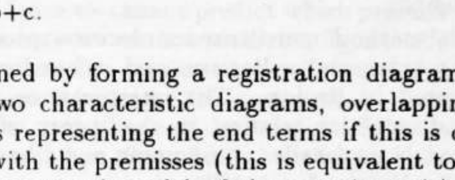

Fig. 4: Registration diagram for the problem "All of the a's are b's, All of the c's are b's", which lacks valid conclusions. Note that the shaded end-term circles overlap, bisecting each other.

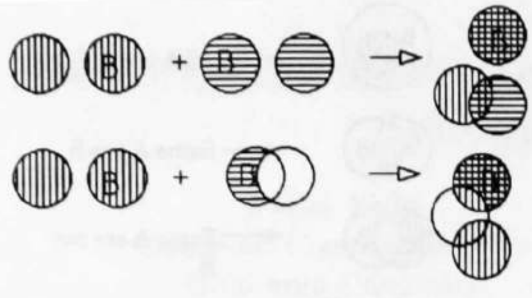

Fig. 5: Registration diagrams for the U-valid problem types, which establish necessary individuals but lack quantified conclusions.

"Some sentient creatures are not Martian"). For universal conclusions with subject X, drop the middle term feature and pick a feature +X such that X corresponds to an unbisected circular shaded region in the registration diagram. Note that although the "double negative" problems corresponding to the diagrams in Fig. 5 establish necessary individuals, they lack quantified conclusions because neither of the end-term features is +. It would be necessary to have another statement type "Some of the not A are not B", which we call type U, to express these conclusions.

This method constitutes a decision procedure for the categorial syllogism, and it has been implemented in Prolog. The interpretation of the method, and its relation to the theory of Mental Models, is detailed in Stenning and Oberlander (1992) and Stenning (1992).

The method can be adapted to provide a range of performance models of syllogisers. Our approach is to adapt the competence model to give a model of correct performance, which permits breakdowns to account for subjects' errors. Erroneous conclusions can be accommodated by assuming that subjects use sub-maximal diagrams, or register diagrams sub-maximally. This is equivalent to assuming subjects fail to consider all possible logical models. We assume the inference process involves the identification of a shaded region in the characteristic diagram for one of the premisses, followed by a test using the information from the other premiss to find out if the shaded region is bisected by registration or not. For optimal performance this process must be iterated for each shaded region, and failure to do so may result in an erroneous "No valid conclusion" response. We call the premiss which provides the shaded region the *source premiss*. We make no claims about the way in which subjects find the source premiss; it could be systematic or random, exhaustive or not. But by hypothesis, production of a valid conclusion entails that the subject has successfully identified the source.

We claim that the Figural Effect described by Johnson-Laird *et al* (Johnson-Laird & Steedman 1978, Johnson-Laird 1983, Johnson-Laird & Bara 1984, Byrne & Johnson-Laird 1991), is a consequence of the assymmetry between the roles of the two premisses in establishing necessary individuals. Basically, the Figural Effect is the tendency for terms to preserve their grammatical status from premisses to conclusion. So in the problem *Some of the Artists are Beekeepers, All of the Beekeepers are Chemists*, both of the conclusions *Some of the Artists are Chemists* and *Some of the Chemists are Artists* are valid, although the vast majority of subjects produce only the former. In figure 1, then, the preferred conclusions are ac ones, in figure 2 they are ca ones, while in the other two figures both types are equally common. Johnson-Laird's account of this effect is touched on briefly below, but is described in detail in the sources cited above.

We assume that the type descriptor is built up incrementally during the reasoning process, so the terms from the source premiss should precede the end term from the other premiss. Given that quantified conclusions will tend to preserve this ordering if possible, response biases toward conclusions with one or other end-term ordering are explicable in terms of which of the premisses is the source. This tends to vary with the figure of the syllogism. Consider the problem AabAbc (see Fig. 3). Although the characteristic diagrams for both the premisses contain shaded areas, only the shaded area from the first, corresponding to the term a, is unbisected by registration. So we predict conclusions in which a is the subject, namely Aac or Iac, but not Ica, although all three are valid. There are three problems in figure 1 that permit valid conclusions with both possible term orders (ac and ca). For two of these problems, the only possible source premiss is Premiss 1, and for the other, both premisses are potential sources (see Fig. 6 below). Therefore the theory predicts figural effects for the first two, but does not predict a figural effect for the third (but see below). Similarly in figure 2, there are three problems with free term order in valid conclusions, two of which have source Premiss 2, and the other has two potential sources. In figures 3 and 4, we can also predict term orderings in individual problems, and the predicted numbers of ac and ca orderings are equal.[1]

Our approach therefore offers an extension of the traditional Eigural Effect, since we can make specific predictions in figures 3 and 4, but it is at present limited to predictions for valid conclusions only. This is because only in these cases is there a principled basis on which to decide what the source premiss is, but in principle the theory should be ex-

[1] Figure 2 problems are equivalent to figure 1 problems when the premiss order is inverted, and in the remaining two figures each problem has an equivalent inverted problem in the same figure.

tensible to handle the Figural Effect in invalid conclusions. To do this, it is necessary to identify the representations used by subjects: then the source premiss can be identified *post hoc*, providing a basis on which to predict term orders. We intend to attempt this work in future.

Although Mental Models theory can at present offer an account of the Figural Effect in invalid conclusions as well as in valid conclusions, the present approach has the advantage that the Figural Effect as we treat it follows from logical rather than physical features of the representations used in inference. Consequently it is not so dependent on the assumption that all subjects follow the same procedure when solving syllogisms, which is rather implausible considering the great differences in experience between different subject groups.

The traditional syllogism task has a "degrees-of-freedom" problem, because for many syllogisms the order of the end terms in valid conclusions is not free to vary, except in I and E conclusions. However, in a task in which subjects are asked to identify the individuals entailed by a pair of premisses, the order of mention of the terms is completely independent of the validity of the conclusion. Additionally, there are valid conclusions for more problems than in the traditional task, since some of the double-negative problems establish individuals but none of the conclusions A, I, E or O. These are the problems which establish 'U-conclusions'. Our theory maintains that the identification of necessary individuals is a prerequisite to drawing a quantified conclusion, so this task is relevant to the traditional task.

With this task, we also need to be able to predict the position of the b term, which can be done by specifying the order in which the terms from the source premiss are mentioned. In the case of universal premisses, each candidate individual is uniquely identified by a single positive term, for example, the minimal model of *All of the a's are b's* is just the set a, all of whose members are implied to be b, so the predicted ordered type descriptor is +a+b. Particular premisses are more troublesome, but for the present we assume that the subject feature precedes the predicate feature. Type descriptors for necessary individuals are then composed of the ordered type descriptor for the candidate individual, followed by the feature corresponding to the remaining end term.

Although the present theory makes predictions about the ordering of end terms in valid quantified conclusions rather similar to those made by Mental Models theory, the predictions of the two theories diverge when we consider the necessary individuals task. According to Mental Models theory, subjects create a mental model in which tokens representing instances of the middle term occur *between* tokens representing instances of the end terms, in a two-dimensional spatial array, and then read off conclusions from one end of the model to the other, so the only predicted orders would presumably be abc and cba. However, the present theory additionally predicts occurrences of the orders bac and bca.

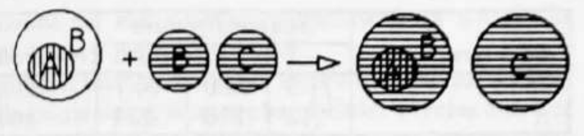

Fig. 6: Registration diagram for the problem "All of the a's are b's, None of the b's are c's". Note that there are two unbisected shaded areas.

Finally, this task allows us to make predictions concerning some of the problems whose source can be either premiss. The problems in question are those which have an A premiss whose subject is an end term, and an E premiss (e.g. AabEbc - Fig. 6 shows the registration diagram). Problems of this type establish 2 individuals, each of which has a different source premiss. According to the theory, each of these individuals supports a different quantified conclusion, since each has only one positive feature corresponding to an end term (e.g. +a+b−c supports Eac and +c−b−c supports Eca). Since we cannot predict which premiss will be source here we cannot predict a figural effect, but we can test for an association between the source and term order, since the source can be determined *post hoc* on the basis of the conclusion that is drawn by the subject. Specifically, answers whose source is the A premiss should have the end term from that premiss before that from the other, and answers whose source is the E premiss should have the opposite end term order.

Method

Design Each subject produced one answer for each of the 64 syllogisms, which were presented in random order.

Subjects 21 subjects, either Psychology undergraduates or postgraduates in Cognitive Science, were each paid £5 to take part.

Materials Each subject received a set of 64 slips of paper, on each of which was printed a different pair of premisses. The vocabulary used was selected from sets of nouns denoting nationalities, professions and interests, for example *None of the musicians are chessplayers. All of the musicians are Italians.* Each vocabulary item appeared in two syllogisms, and two different random assignments of vocabulary to syllogisms were used.

Procedure Subjects were instructed to imagine that the premisses described a group of people at a party, and to decide whether any kind of person who could be described with certainty, in terms of

End term order	Source Premiss 1	2	Both	Overall
ac	77.9	28.0	65.7	55.8
N	213	218	134	565

Table 2: Percentages of correct valid conclusions with end-term ordering ac in problems with source premiss first, second or either.

	Figure 1	2	3	4	Overall
Predicted	86.2	79.5	85.2	62.2	75.4
N	138	122	88	217	565

Table 3: Percentages of predicted term orderings in correct responses to problems with valid individual conclusions in each figure.

either positive or negative values of all three features, had to be present in the room. They were instructed to assume that some people corresponding to each of the three terms existed. They were asked to describe the individual on the slip of paper, or if there was no individual, to write "No valid conclusion" Subjects worked alone in quiet surroundings, and were given as much time as they needed to finish all the problems.

Results

Table 2 shows how the order in which the end terms are mentioned, in type descriptors for correct answers to problems with valid conclusions, varies with source premiss. The variables are strongly associated ($\chi^2(2) = 116.0$, $p < 0.0001$), such that a tends to precede c when the first premiss is the source, and c tends to precede a when the second premiss is the source. Note that there is some overall tendency for a to be mentioned before c, particularly in problems where either premiss can function as the source.

Table 3 shows the percentages of correct responses to problems with valid conclusions which occur in the predicted orders in each figure of the syllogism. Overall, 75.4% of responses were as predicted, and there was a majority of predicted responses in each figure. However, the effect is not entirely independent of figure ($\chi^2(3) = 34.77$, $p < .0001$), the main divergence from the overall trend being in figure 4, where a substantial minority of responses occurred in unpredicted orders.

Table 4 shows the frequencies of responses having each of the possible term orderings. It is clear that, against the presumed predictions of Mental Models theory, responses with bac and bca orders actually outnumber those with abc and cba orders. As predicted by both accounts, there are very few cases in which the b term occurs last.

	Order abc	cba	bac	bca	cab	acb
N	194	148	236	156	31	19
Total	342		392		50	

Table 4: Frequencies of responses with each possible term ordering (total N=784).

End-term order	Source Premiss A	E	Total
AE	64	0	64
EA	4	10	14
Total	68	10	78

Table 5: Association between end-term order and source premiss for valid conclusions to problems which establish two individuals. End-term order is said to be AE if the end term from the A premiss precedes that from the E premiss.

Finally, Table 5 shows the relation between end-term order and source premiss for all of the the problems which establish two individuals (e.g. Fig. 6). On the basis of the response, we can determine whether the A or E premiss is the source, and we predict that the end-term from the source premiss should precede the end-term from the other. As the Table shows, there are more responses with the A premiss as source than with the E premiss as source, and the end-term order is strongly predicted by the source premiss (*Yates'* $\chi^2(1) = 46.24$, $p < .0001$).

Discussion

The results of the experiment show that the source premiss for a problem is strongly associated with the order of mention of the end terms, in both determinate and indeterminate problems. The three-term order predictions are also confirmed, and with the exception of the anomalous result for figure 4 (discussed below), the effect is uniform across figures, so the figural effect is explained by the distribution of different problem types among the figures. So our model can offer a uniform account of order effects in both this task and the traditional syllogism task. Although Mental Models theory handles the phenomena of the traditional task well, it fails to predict the large number of responses in this task in which the b term occurs first, so it cannot at present offer a uniform account of the two tasks.

Mental Models theory, of course, purports to account for other aspects of performance in the traditional task. Johnson-Laird (e.g. 1983) has argued that the number of mental models which need to be considered to solve a problem predicts its empirical difficulty quite well. However, Ardin (1991) has shown that in this experiment's data, although one-

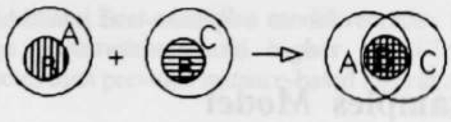

Fig. 7: Registration diagram for the problem "All of the b's are a's, Some of the b's are c's".

model problems are easier than the others, there is no significant difference between two- and three-model problems. Our theory does not rely on the construction of different models, so it cannot account for differences in difficulty in such terms, but it may be that difficulty can be predicted by the number of candidate individuals which have to be considered. This has yet to be investigated.

The results of the experiment showed that a substantial minority of correct responses to figure 4 problems did not occur in the predicted orders. There are indications that this is due to another strategy which can occur when the non-source premiss is of type A. Subjects sometimes produce the terms from this premiss before those from the "true" source premiss, but only under special conditions, when the shaded region from the A premiss is bisected by registration, but one of the halves is itself an unbisected shaded region, and so represents a necessary individual. Ordinarily, we would expect subjects who have detected bisection of a region either to give a "no valid conclusion" response, or start afresh with a shaded region from the other premiss to find the unbisected shaded region, but in this case it appears that they can detect the critical region on the first pass, and so produce the terms in the order which would be predicted if the A premiss was the source. An example is Abalbc (see Fig. 7). There are similar problems in all the figures, but most are in figure 4, where the effect is most pronounced. The tendency for subjects to consider the A premiss first, or the "A-effect", has previously been noted by Lee (1987).

The results of the analysis of the Fig. 6-type problems also suggest the presence of an A-effect, insofar as most of the valid conclusions to these problems are only accessible from the A source rather than the E source. This, along with the core theory, can explain the figural effect for these problems, as follows. For the problem AabEbc, the A-source individual conclusion is +a+b−c, which supports only the quantified conclusion Eac, since the positive end-term feature is a, which must therefore form the subject of the conclusion, so the conclusion is figural. Similar arguments hold for the figure 2 problem, as well as the two in figure 3, one of which supports Eac and one of which supports Eca. Subjects' preferences concerning which premiss to consider first are independent of the central claims of this paper, but it appears that augmenting the theory by including the A-effect would successfully account for some figural phenomena on which the core theory remains neutral.

In conclusion, it is clear that a method for solving syllogisms using isomorphs of Euler Circles can not only avoid large numbers of representations, but can also explain some classic psychological results using minimal auxiliary assumptions. We hope soon to produce evidence that the theory can also account for the Figural Effect in invalid conclusions.

Acknowledgements

Thanks to Cath Ardin and Morten Christiansen for work on the preparation and running of the experiment, and to Jon Oberlander for preparing some of the diagrams.

References

Ardin, C. 1991. Mental Representations Underlying Syllogistic Reasoning. Ph.D Thesis, Centre for Cognitive Science, Edinburgh University.

Byrne, R. and Johnson-Laird, P.N. 1991. *Deduction*. Hove: LEA.

Erickson, J.R. 1974. A set analysis theory of behaviour in formal syllogistic reasoning tasks. In R.Solso (ed) *Theories in cognitive psychology: the Loyola Symposium*. Potomac: LEA.

Guyote, M.J. and Sternberg, R.J. 1981. A transitive-chain theory of syllogistic reasoning. *Cognitive Psychology* 13:461-525.

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

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

Johnson-Laird, P.N. and Steedman, M.J. 1978. The psychology of syllogisms. *Cognitive Psychology* 3:395-400.

Lee, J. 1987. Metalogic and the psychology of reasoning. Ph.D Thesis, Centre for Cognitive Science, Edinburgh University.

Stenning, K. 1992. Spatial inclusion as an analogy for set membership: a case study of analogy at work. To appear in Holyoak, K. and Barnden, J. (eds) *Analogical Connections*.

Stenning, K. and Oberlander, J. 1992. A cognitive theory of graphical and linguistic reasoning: logic and implementation. Research Paper RP20, H.C.R.C., Edinburgh University.

Woodworth, R.S. and Sells, S.B. 1935. An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology* 18:451-460.