Image and language in human reasoning: a syllogistic illustration

Keith Stenning & Peter Yule

A version of this paper has been submitted
to *Cognitive Psychology*, Oct 1995.

Address for correspondence:

Professor Keith Stenning
Human Communication Research Centre
Edinburgh University
2, Buccleuch Place
Edinburgh EH8 9LW
Scotland
Tel.: +44 31 650 4444
Email: keith@cogsci.ed.ac.uk
FAX: +44 31 650 4587

Abstract

Cognitive theories of human reasoning can benefit by distinguishing logics from implementations. An semantic analysis of the syllogism reveals what is common between apparently unrelated methods of reasoning and how such an analysis can guide experimental investigation.

Existing accounts of syllogistic reasoning oppose rule-based linguistic to model-based methods. Stenning & Oberlander (1995) show that the latter are isomorphic to well-known graphical methods correctly interpreted. We here extend these results by showing that equivalent sentential implementations exist, thus revealing that all these approaches are members of a family of abstract *individual identification algorithms* (IIAs) variously implemented in diagrams or sentences.

We investigate whether subjects employ the IIA through an experiment using a novel reasoning task. Modelling of this data shows that behaviour is best explained at the abstract level of the IIA with consequences for the semantic and the empirical issues facing cognitive models of reasoning and mental experience.

1 Introduction

Understanding the relation between logic and human performance is fundamental for cognitive theories in many domains—reasoning, decision making, planning, communication. The distinction between logics and implementations is commonplace in computer science, and the relation between them is well known to be many-levelled and indirect. If this is so for machines, we should expect it to be more so for human beings. The aim of this paper is to illustrate, for cognitive theories of human reasoning, the benefits of distinguishing carefully between abstract logics and their implementations in analysing mental processes.

Cognitive theories of mental process are often distinguished by the nature of the representations they invoke in the implementation of their computational processes. A particularly prevalent contrast is between linguistic and imagistic representations. Arguments about distinguishing these two types of mental representation—the 'Imagery Debate'—have a long history (see Block (1981) for a review). Alongside the enduring impulse of psychologists to contrast them, arguably driven by personal phenomenology, runs a current of comment which denies that available evidence can distinguish them (e.g. Pylyshyn (1973), or even the theoretical possibility of distinguishing them (Anderson 1978).

As has often been pointed out, representations (whether pictures, sentences, or images) only play their explanatory role when embedded in *systems*—systems of semantic interpretation and systems of inferential transformation. Indeed the very term *propositional* which has been adopted in cognitive science to indicate linguistic representation, has its recent origins in the logician's need for a term which is *invariant* with regard to representation. Propositions, by definition, may as well be expressed by pictures as by sentences. The best understanding of the arguments against distinguishability of imagery and linguistic representations is as a *sceptical* argument. Language is ineffable (or so the argument goes)—there is nothing that cannot in principle be represented in language in general. So to merely claim that representation is propositional is to make no contingent claim. The purpose of this argument is to make it clear that it is incumbent on proponents of imagery to propose particular accounts of the actual systems in operation to say how they are not universally expressive. Characterising particular systems is a matter of saying what they *cannot* express.

Some recent theories of the cognitive efficacy of different modalities of *external* representations have taken up the challenge of analysing the computational differences between diagrams and texts. Larkin & Simon (1987) have proposed that when graphics is 'good' it is because graphical representations facilitate certain computations—e.g. parallel search. This view proposes that a graphical and a linguistic representation may be *informationally* identical but *computationally* distinct. Stenning & Oberlander (1995) emphasise the distinction between *token* representations and *systems* of representation in applying the concept of expressiveness. They propose that when graphics is 'good' it is because graphical *systems* are inexpressive relative to languages. Although *token* linguistic and graphic representations may be informationally identical, the linguistic ones are generally drawn from systems which are much more expressive (ie have tokens for which there are no graphical equivalents) and that this is the reason why a general sentential processor cannot be so efficient. Specialised sentential processors can be just as efficient as graphical ones for weak sentential fragments. But it may be harder for a naive reasoner to exploit the weakness of sentential fragments than for graphical fragments because of the availability of expressed constraints.

All of these arguments point up the importance of distinguishing between different levels of abstraction over algorithms in making claims about types of representation, and in con-

necting those claims with behavioural and phenomenological evidence. What this paper sets out to do is to explore the consequences for cognitive theory of a thorough semantic analysis at several levels of abstraction in one example domain—syllogistic reasoning. The major distinction between psychological models of mental representation and process in this field is between sentential and model-based theories. Stenning & Oberlander (1995) have clarified the 'graphical' nature of the model-based theories, in showing that the mental models method (Johnson-Laird 1983) is strongly equivalent to a rational reconstruction of the Euler's (1772) (and Leibniz' 1666/1966) method. In this paper, we show that the graphical algorithm is also isomorphic to a natural presentation of a sentential natural deduction method. This demonstration reveals important computational correspondences between sentential and imagistic representations; relates syllogistic reasoning to more routine discourse processing; suggests novel ways of gaining behavioural evidence about the mental process of syllogistic reasoning; and raises fundamental questions about the relation between computation and phenomenological experience.

Specifically, we will show that a careful study of the semantic structure of the syllogistic fragment reveals a model-theoretic property we call *case-identifiability* (CI)—all and only syllogisms which entail the existence of a maximally specified type of individual (specified in terms of all three properties A, B, and C), have a valid conclusion. We adopt the term CI to suggest this metalogical property that all conclusions are valid in virtue of an identifiable 'case'. Because the syllogism is case-identifiable, there is one abstract family of algorithms (which we call *individual identification algorithms* (IIA)) which characterise a critical type of individual, and then draw conclusions by abstracting away from this characterisation. Whilst there may be many other algorithms for syllogism solution, we claim that all extant psychological theories of how people solve syllogisms fall into this abstract class.

This demonstration raises two questions. Firstly, do people solve the task in terms of some implementation(s) of the abstract individual identification algorithm for solving syllogisms, or have all the theories which have agreed on this point been wrong? Secondly, what sort of evidence could distinguish between linguistic and imagistic sub-families of the abstract algorithm? If imagistic and linguistic algorithms have not been specified in enough detail to distinguish them, how could they be distinguished and what sort of psychological evidence would decide between these alternatives?

In this paper we will concentrate on the first question. Our semantic characterisation suggests novel ways of collecting fresh empirical evidence that people do use the individual identification algorithm, evidence that does distinguish it from other approaches. We present such evidence along with statistical models which uncover new empirical regularities hitherto unnoticed in the conventional data. We relate those regularities to features of our semantic analysis. This leaves the second question about what could distinguish between linguistic and imagistic *mental* implementations of the algorithm. Merely showing that well specified external graphical and linguistic algorthims are isomorphic, and that existing psychological data does not distinguish their mental implementations, does not demonstrate that they are in principle indistinguishable. We return to this second class of issues in the general discussion.

The plan of the paper is as follows. We begin with a brief description of the several levels of abstraction at which logic operates. Then beginning at the most abstract model-theoretic level, we illustrate the syllogism's property of case identifiability. This property is then shown to enable the deployment of the individual identification class of algorithms which is then presented in the form of a `prolog` implementation. Stenning & Oberlander's (1995) reconstruction of Euler's algorithm (redescribed here for convenience) is an example

of a graphical implementation of this general class of algorithms. In order to compare this graphical treatment with a linguistic one, we then present a natural deduction algorithm, and show that there is a direct but hidden correspondence between graphical and natural deduction implementations of the general algorithm.

Turning to the empirical consequences of our analysis, the experiment uses a novel individual identification task to gather data about the sequencing of properties in subjects' mental reasoning to test whether or not the individual identification algorithm is in fact used by subjects performing 'mental reasoning'. A model of this data is then presented which reveals new generalisations about subjects' performance. The Standard Task condition of the experiment is designed to show that the new results are not an artefact of the novel task but can be found in the conventional task.

The general discussion interprets the implications of characterising human syllogistic reasoning as employing this general class of algorithm, and of the equivalence of its graphical and linguistic implementations. Finally we return to the question about what does set imagistic and linguistic methods apart, and what sorts of psychological evidence might distinguish them.

2 The logical structure of the syllogism

2.1 Levels of logical analysis

Within logic, there are several levels of theory which bear different relations to processes of deduction. Logics are, most fundamentally, consequence relations—abstract mappings between sets of premisses and valid conclusions. Consequence is a *semantic* relation. If a statement is true in any model of a set of premisses, that statement is a valid consequence of those premisses. Model theory, the part of logic which studies ranges of models in which statements are true and consequents valid, does *not* treat of processes of deduction.

At the next, less abstract level, proof-theory, the study of reasoning in terms of the syntax of sentences (or the *form* of non-sentential representations) does provide descriptions of sequences of reasoning in terms of transitions from sentences to sentences. While proof-theory provides rules for transforming sentences, and hence descriptions of chains of deduction, it provides no guidance as to which transformations to apply to premisses in order to reach a target conclusion. It still describes infinite abstract mappings between premisses and conclusions. These mappings have more structure but they are still static mathematical abstractions. Metalogic consists of proofs about the coincidence of semantic and syntactic properties of logics. *Sound* proof systems prescribe syntactic operations which preserve truth. *Complete* proof systems provide syntactic operations which can reach all the semantic consequences that any consistent system should reach.

At the next less abstract level, theorem provers provide guidance for syntactic proof processes. They guide a proof towards a target conclusion. Sometimes, in *decidable* systems, this can be done in an algorithmic fashion which guarantees reaching a target conclusion if it is valid, or its negation if it is invalid. Sometimes it is possible to prove metalogically that no such theorem prover can exist and the best that can be done is to provide heuristic theorem proving guidance. It is at the level of theorem proving that logic begins to describe *processes* of deduction. These processes can be described in ever increasing detail right down to physical instantiation of the computational processes involved. Marr's description of levels of abstraction has often lead to the inference that there are just three. Foster (1992) pro-

Number	1	2	3	4
Term	a-b	b-a	a-b	b-a
Arrangement	b-c	c-b	c-b	b-c

Figure 1: The four figures of the syllogism, with their conventional numbering.

vides a framework for describing abstraction which is more compatible with the possibility of indefinite numbers of levels.

The syllogism is a small decidable fragment of logic which is easily supplied with decision procedures. It is perhaps for this reason that psychologists often treat logical accounts as *mechanisms* of reasoning. Small though the fragment may be, we will show that distinguishing levels of abstraction is important to psychological understanding.

2.2 Model Theory

Syntax Our classification of syllogisms follows that proposed by Johnson-Laird (1983). There are 4 *moods* or premiss types, with the quantifiers (and conventional mnemonics) "all" (A), "some" (I), "none" (E) and "some...not" (O). A syllogism consists of two premisses, each of which relate two terms, one of which (the middle term) occurs in both premisses, while the other two (the end terms) each occur in only one premiss. There are four possible arrangements of terms in the two premisses, known as *figures*, as shown in Figure 1. Since each premiss can be in one of four moods, and each premiss pair can have one of four figures, there are $4 \times 4 \times 4 = 64$ different syllogisms.

Semantics The syllogism can be viewed as a fragment of the monadic (one-placed) predicate calculus. Because the monadic calculus has no identity relation, its semantics concerns *types* of individuals defined by combinations of non-relational properties—it is, therefore, never of semantic significance how many of a type of individual exist. *Maximal* types are described by combinations of all three predicates of a syllogism. There are eight such types: ABC, AB¬C, A¬BC,A¬B¬C, ¬ABC, ¬AB¬C, ¬A¬BC, and ¬A¬B¬C. *Interpretations* of the syllogism consist of sets of these maximal types of individual. Conventionally, the syllogism is interpreted under the assumption that none of the three sets A, B and C are empty. Subjects naturally adopt this assumption (as witnessed for example by their readiness in the task situation to conclude from *All A are B* that *Some A are B*). This 'no-empty-sets' axiom reduces the number of possible models somewhat.

Inferential Structure Of the 64 syllogisms 27 have conventionally valid conclusions which can be formulated by applying one of the four quantifiers to the two end terms. The remaining 37 syllogisms do not allow any valid conclusions within the vocabulary of conclusions thus specified, though as we will see, there is a further group which allow valid conclusions if a further quantifier is available.

Examination of the valid syllogisms reveals that all syllogisms which have valid conclusions entail the existence of at least one *maximal* type of individual. So, for example, *Some B are ¬A. All B are C.* entails that there are individuals of the maximal type ¬ABC (and therefore that *Some C are not A*). There is obviously a direct relation between the entailment of the existence of such maximal types and the entailment of such existential conclusions. What

concerns us here is that the existence of the maximal individual is semantically richer than the conventional existential conclusion. Table 11 in Appendix A.1 shows the necessary individuals for each of the 64 premiss pairs; some readers may prefer Figure 12 in Appendix A.3 for a more perspicuous graphical presentation of this information.

Considering the converse of this generalisation is revealing. It turns out that for all but one small group of syllogisms, if they entail the existence of a maximal type, they also have a valid conventional conclusion. The exceptions are a group of syllogisms with two negative premisses (one of the main generalisations of Aristotle's meta-theory of the syllogism was that two negative premisses had no valid conclusion). For example, *No A are B. No B are C* establishes that there are $\neg AB\neg C$ individuals. The quantificational apparatus of the conventional syllogism will not allow expression of the conclusion *Some* $\neg A$ *are* $\neg C$ (we will refer to these as 'U' conclusions following the common notation of the four conventional quantifiers as A=all, E=none, I=some and O=some not). There are good historical reasons why Aristotle did not include these conclusions in his theory which turn on his lack of a distinction between local and universal domains.

These model-theoretic properties of the syllogism are highly unusual. Even other small finite logical fragments are not generally case-identifiable. For example, disjunctive syllogisms such as the argument *Everything is A or B. Everything is A or C. Nothing is A. Therefore everything is B or C.* are not case identifiable. These premisses are satisfied by the model consisting of the following set of types: $\{\neg ABC, \neg AB\neg C, \neg A\neg BC\}$, and any of its subsets, but none of the types are entailed to exist by the premisses. These disjunctive arguments are not case-identifiable.

Case-identifiability, a model-theoretic property of the syllogism, plays a crucial role in determining what representations and algorithms can be employed on this fragment of logic. In particular, the graphical methods described below rely on the fact there is no need to represent *partially* specified individuals in the diagrammatic representation of the conjunction of two premisses. Having an *efficient* graphical strategy which need construct only one diagram per syllogism is further dependent on the even stronger property that the *maximal* model for any syllogism is unique, and consists of all the individuals consistent with the premisses (i.e. there are no contingencies between the presence/absence of types in a model, unlike, for example the disjunctive argument example above). The natural deduction fomulation of the abstract algorithm given below is also dependent on the case-identifiability property.

2.3 Solving syllogisms by identifying critical individuals

Our purpose in this section is to describe an abstract algorithm for syllogistic reasoning, the IIA. Initially, this algorithm is specified in model-theoretic terms, with a `prolog` implementation. We then describe both a graphical and a sentential formulation (Euler's Circles and a natural deduction method) of this abstract algorithm. Our purpose is to reveal commonalities between what have been taken to be distinct processes defining opposing psychological theories of reasoning.

2.3.1 The Abstract Individual Identification Algorithm

Premiss representation The meaning of premisses is represented in terms of sets of individuals which satisfy them. Model-theoretically, each premiss type is satisfiable by any of several different sets of individual types. However, the present method makes use of only two

of these sets for each premiss, these being the set of possible individuals whose existence is consistent with the premiss, known as the *maximal model (Max)*, and the set of individuals whose existence is entailed by the premiss, known as the *minimal model (Min)*. These are respectively the largest and smallest sets of types compatible with the premiss. The representation can thus be characterised as the ordered pair $< Max, Min >$. Since $Min \subseteq Max$, it is feasible to represent each premiss type using a single integrated representation with individuals marked as necessary or possible (cf. Mental Models).

Types of individuals are notated by feature structures, where features are constructed from the terms appearing in the premisses of the syllogism, prefixed by "+" or "−", which indicate whether the individual concerned is or is not a member of the set denoted by the term. For example "+P+Q", denotes a type of individual which is P and is Q, i.e. $(\exists x)(Px \& Qx)$. Similarly, "+Q−R" denotes a type of individual which is Q and is not R, i.e. $(\exists x)(Qx \& \neg Rx)$. Feature structures are unordered owing to the commutativity and associativity of conjunction.

Case Identification Case identification depends on the integration of information from both premisses. The aim is to identify the minimal model of the premiss pair, the set of feature structures which are specified with respect to all three terms in the problem, which denote those individuals whose existence is entailed by the premisses taken together. This can be achieved using unification, an operation which takes two feature structures and returns a third which is their set-union, provided the two do not contain different values for any of their members. So, for example, the unification of +A+B with +B−C is +A+B−C, but the unification of +A+B with −B−C is undefined, since the two do not have the same values for the B feature.

The set of all unifications of each member of one maximal model with each member of the other (the Cartesian product of the two sets) is the maximal model of the premiss pair, containing all types compatible with the premisses. Finding the minimal model turns on unifying members of the minimal model of one premiss (*critical individuals*) with members of the maximal model of the other premiss. We attempt to unify each critical individual i from one of the premisses (P) with all the members j of the maximal model of the other premiss (Q). If only one such unification is possible for some i, then the feature structure which is formed must denote a member of the minimal model of the premiss pair, because it is the only possible completion of a type which itself must exist. For example, supposing i is +A+B and $j \in \{$+B+C,−B+C,−B−C$\}$ (the maximal model of "All the Bs are Cs"), the only possible unification is +A+B+C. We know that +A+B must exist, and since +A+B−C is impossible, it follows that +A+B+C must exist. We call the premiss P, which supplies the critical individual i, the *source premiss*, and since the other premiss Q is always a universal, we will refer to it as the *conditional premiss*—this distinction has consequences in the empirical investigation in Section 3 below.

If every critical individual from both premisses has been tested but no necessary individuals have been found, then the syllogism has no valid conclusion relating the end terms.

Drawing Quantified Conclusions Feature structures denoting necessary individuals form the basis for quantified conclusions. Since these relate only the end terms, the middle term feature can be deleted from the feature structure. Of the remaining features, if one is positive, a conclusion with that term as subject is warranted, since the grammar of conclusions does not

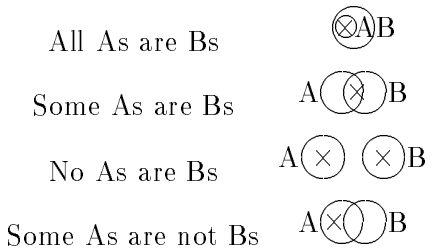

Figure 2: Characteristic diagrams for each premiss type. Regions representing necessary individuals are marked with a $\times$.

permit negation in subject position[1]. The sign of the remaining feature determines whether the conclusion is positive or negative. This is all that is required for a particular conclusion, so that given e.g. +A+B−C, we delete +B giving +A−C, the only positive term is +A so A is the subject, and since the remaining term −C is negative, the conclusion is "Some of the As are not Cs". For universal conclusions, a further test is required. This can be characterised as follows: take the subject feature, and unify it with all members of the maximal model for the premiss pair. If only one such unification results, a universal conclusion is warranted, so in the example just given, provided the test was positive, we could conclude "None of the As are Cs". This test is equivalent in graphical terms to checking that the circle named by the subject term is not bisected in the registration diagram.

In Appendix A.2 we present a `prolog` implementation of the method. This set-theoretic `prolog` implementation is the most neutral and most directly related to the syllogistic consequence relation. We now show how a graphical treatment can implement the same abstract algorithm.

2.3.2 The Graphical Method

The graphical algorithm was presented in Stenning & Oberlander (1995) where it is shown that Johnson-Laird's 'mental models' method is an alternative implementation of the same abstract algorithm. We restate it here for the reader's convenience.

Premiss representation In the graphical implementation of the method, individual types are represented as regions in a diagram delineated by circles. The maximal model of a premiss is represented as a diagram containing two circles, one for each of the terms. Regions in the diagram correspond to individual types which are compatible with the premiss (the maximal model), while the minimal model is represented by marking with an "$\times$" the regions of this diagram which correspond to necessary individuals. Figure 2 shows the characteristic diagram for each premiss type. The circles are assumed to be of variable size and position. Only their topological relationships (patterns of containment, overlap and exclusion) are significant. This premiss representation is a direct translation of the set-theoretic relations into graphical terms.

[1]U-conclusions, which have only negative end terms, are barred by this condition - see Stenning & Oberlander (1994)

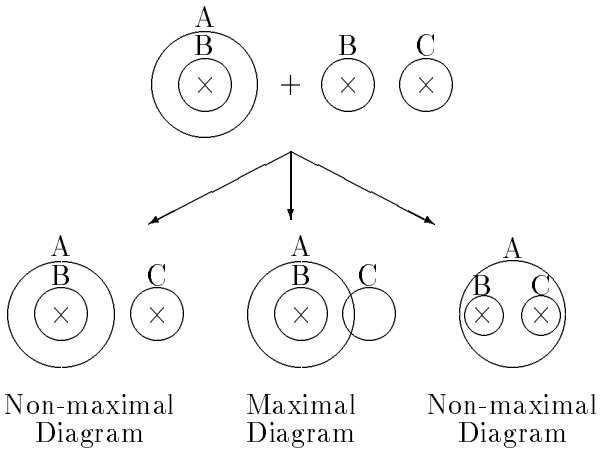

Figure 3: Constructing a registration diagram for the example "All of the Bs are As, None of the Bs are Cs". While any of the three possible registration diagrams is consistent with the premisses, the central one is chosen since it has the largest number of regions.

Case Identification A new diagram is constructed by 'registering' the two characteristic diagrams by superimposing the B circles. There may be more than one way of doing this consistent with the premisses—making it a "multiple models" problem (Johnson-Laird & Bara (1984)—but in the present normative method only one way is chosen, namely the one which creates the maximum number of sub-regions in the diagram. This is achieved by overlapping the circles representing the end terms if this can be done while remaining consistent with the premisses. The result is the *registration diagram*, which represents the maximal model of the premiss pair. Figure 3 shows an example; in this case the centre registration diagram is the maximal one. Now each marked region deriving from the original characteristic diagrams will either have been bisected by the third circle or not. The marked region represents i, a critical individual, so it can be called a *critical region*. If the critical region is bisected, then the mark is removed. The attempt to overlap the end-term circles ensures that all possible unifications of i with each j have been attempted, so if a critical region is not bisected, and consequently remains marked in the registration diagram, only one unification is possible, so it represents a critical individual. In the example it is the individual +A+B+C, and since the marked region in the final diagram is the B circle, which appears marked in both the characteristic diagrams, both premisses are source premisses in this case.

Registration diagrams can form the basis for a logical classification of problem types. The labelling of the end-term circles is immaterial, since premisses can be reordered without affecting the logic of the problem, so the classification abstracts over such differences. Similarly, the lack of topological distinctness between representations of convertible premisses (e.g. *No As are Bs* $\leftrightarrow$ *No Bs are As*) means that the classification abstracts over these differences too. Figure 12, in Appendix A.3, shows the resulting set of 21 distinct registration diagrams, for problems with and problems without valid conclusions. It is useful to list positive and negative syllogisms separately.

Note that this graphical algorithm which remains faithful to Euler's practice, is a deterministic procedure for syllogistic reasoning. It does not suffer from the combinatorial explosion of cases which some misinterpretations of Euler's method exhibit.

All A are B	$A \rightarrow B$
Some A are B	$A\&B$
No A are B	$A \rightarrow \neg B$
Some A are not B	$A\&\neg B$

Figure 4: Propositional representations for each premiss type.

2.3.3 A sentential approach to syllogisms

The syllogism is a logic with quantifiers, and so is normally formulated as a fragment of monadic predicate logic. However, we present here an alternative formulation in purely propositional terms. For the purposes of understanding the relation between linguistic and graphical formulations of solution algorithms, this propositional formulation is more perspicuous because it shears off quantifiers and their variables laying bare the critical propositional operations. Under this interpretation, syllogistic premisses are encoded into propositional calculus according to the scheme in Figure 4.

&-elimination and introduction, $\rightarrow$ elimination, (otherwise known as *modus ponens* (MP)), double negation, and *modus tollens* (MT) are sufficient to draw all valid syllogistic conclusions.

Within this propositional system, we can conceive of the business of drawing syllogistic conclusions as the process of building up descriptions of critical individuals. The descriptions built up are thought of as assertions about single critical individuals. The syllogism is normally developed, and certainly conceived of by naive subjects, as interpreted with a no-empty-sets axiom which asserts that there are some As, Bs and Cs. In our propositional formulation, we think of the no-empty-sets axiom as providing a choice of starting points in the process of building up a description of a single critical individual. We are permitted to assume just one of A, B, or C as the foundation of the description[2].

For example, the syllogism *All A are B. No C are B* would be represented as: $A \rightarrow B$, $C \rightarrow \neg B$. We can proceed by assuming A. Then we can derive B from the first premiss by MP. From B together with the second premiss, we derive $\neg C$ by MT. We can now construct the specification of a critical individual by &-introduction—$AB\neg C$. Having characterised a critical individual, conventional existential syllogistic conclusions can then be drawn by eliminating the middle term by &-elimination. Figure 13 in the Appendix A.4 presents the propositional representation, the initial assumption, and the possible rule used for each valid syllogism.

Figure 5 presents an algorithm based on these methods which derives all and only valid conclusions for syllogisms. The algorithm divides into three parts: identifying the source premiss as the foundation of the ID; completing the ID with information from the conditional premiss by applying MP or MT; and drawing abstract conclusions from the completed ID.

In identifying the source premiss we first seek unique existential premisses, and failing that, unique end-term subjects of universal premisses. Failure to find either (after some possible grammatical reformulation of 'No' premisses) leads to 'no valid conclusion' responses. The second part of the algorithm combines the b conjunct of the source premiss with the conditional premiss as justification of an application of either MP or MT, to conclude the third conjunct of the ID. Finally the third part abstracts conclusions from the ID by eliminating the b conjunct and introducing a quantifier. This abstract description makes it clear that

[2]It is this focus on a single individual which permits the propositional formulation.

this algorithm is a member of the IIA.

2.3.4 Graphical and sentential implementations compared

Although the sentential representations employed in this algorithm are quite different from the Euler diagrams, the two processes both consist of a phase of identification of critical individuals, followed by an abstraction of conclusions. Using this abstract correspondence we can seek detailed correspondences between the representations and operations of the two implementations.

At the basic level of premiss representation, the outlines of the correspondences are clear. Premisses with intersecting lines are existential. Premisses without intersecting lines are universal. Premisses with all $\times$s inside the middle term circles are positive. Premisses with any $\times$s outside the middle term circles are negative. The subject term of the sentential form always corresponds to an $\times$-marked minimal region.

But there is a general difference between the algorithms as they have been stated thus far. The graphical algorithm completes the unification procedure fully determining which subregions appear in the final diagram before reading off any conclusion. The sentential algorithm is focussed on the construction of the ID to the exclusion of other types. In graphical terms, it is focussed on a single $\times$-marked region. It determines whether or not there will be such a region, and to what maximal type it corresponds before deciding any issues related to other regions in the final diagram. Indeed, this is just the intuitive contrast between model-based and linguistic algorithms. But it is not hard to see that this difference can be more apparent than substantial.

Figure 6 provides a graphical expression of the critical sentential rules *modus ponens* and *modus tollens*. The two premiss diagrams with b minimal represent the only two premisses in which b is an antecedent of a conditional, and therefore the only two premisses which enter into modus ponens. On the left are the five diagrams in which the b circle contains an $\times$, corresponding to the five premisses which entail b, the other premiss required for an application of modus ponens. These five are all the diagrams in which b contains an $\times$. Notice that the results of registering each with the diagram representing the conditional always preserves the $\times$ originating in the source premiss. Its inclusion in, and exclusion from, the end-term circles is continued in the final diagram.

The single diagram representing a conditional with a b term irreducably in consequent position appears as the conditional component of an application of modus tollens[3]. The only diagram with no $\times$ in b is the only premiss qualifying as being the negation of the consequent of the conditional, the condition for applying modus tollens to a conditional with a positive consequent. The resulting registration diagram preserves the $\times$ outside b. Its inclusion in, and exclusion from, the end-term circles is continued in the final diagram. The resulting 10 types of registration diagram on the right correspond to all the registration diagrams licencing valid conclusions.

This correspondence of graphical operations to sentential rules shows how an algorithm working over graphical representations does not have to fully compute both maximal and minimal models in order to construct the graphical specification of the critical individual. To illustrate it is necessary to consider a multiple model syllogism such as *All B are A. Some B*

[3]Although universal negatives can be represented as $A \rightarrow \neg B$ they also give rise to $B \rightarrow \neg A$.So for each application of MP to a universal negative conditional, there is a corresponding application of MT which has the same result.

1. Seek a unique existential premiss:
   (a) If there are two, then respond NVC.
   (b) If there are none, then go to 2.
   (c) If there is a unique one, make it the source premiss and go to 3
2. Seek a unique universal-premiss end-term subject:
   (a) If there are none, choose an arbitrary source premiss. Go to 3.
   (b) If there are two, reconfigure a 'No' premiss. On failing, conclude NVC.
   (c) If there is now a unique one, choose its premiss as the source premiss and go to 3
3. If the source premiss is existential, then take its two terms as the first two clauses of the ID. If the source premiss is universal, assume its antecedent. Apply MP and conjoin the consequent to the antecedent to make the first two clauses of the ID.
4. Compare middle terms:
   (a) If a source middle term matches (with regard to negation) the antecedent middle term of the conditional premiss, apply MP, and conjoin consequent term to ID. Go to 5.
   (b) If the source middle term mismatches (with regard to negation) with the conditional consequent middle term, apply MT to the conditional premiss, and conjoin consequent term to ID. Go to 5.
   (c) ELSE conclude NVC and quit.
5. ID is now complete.
6. Draw Abstract conclusion from ID:
   (a) Delete B conjunct from ID. Quantify existentially for an existential conclusion (reordering any positive conjunct into subject position).
   (b) If clause 2c was satisfied, then there is a universal conclusion with the source premiss end-term as subject.

Figure 5: The 'sentential' algorithm based on the process of constructing an *individual description* (ID) by conjoining terms for each predicate or its negation ('no valid conclusion' is abbreviated NVC)

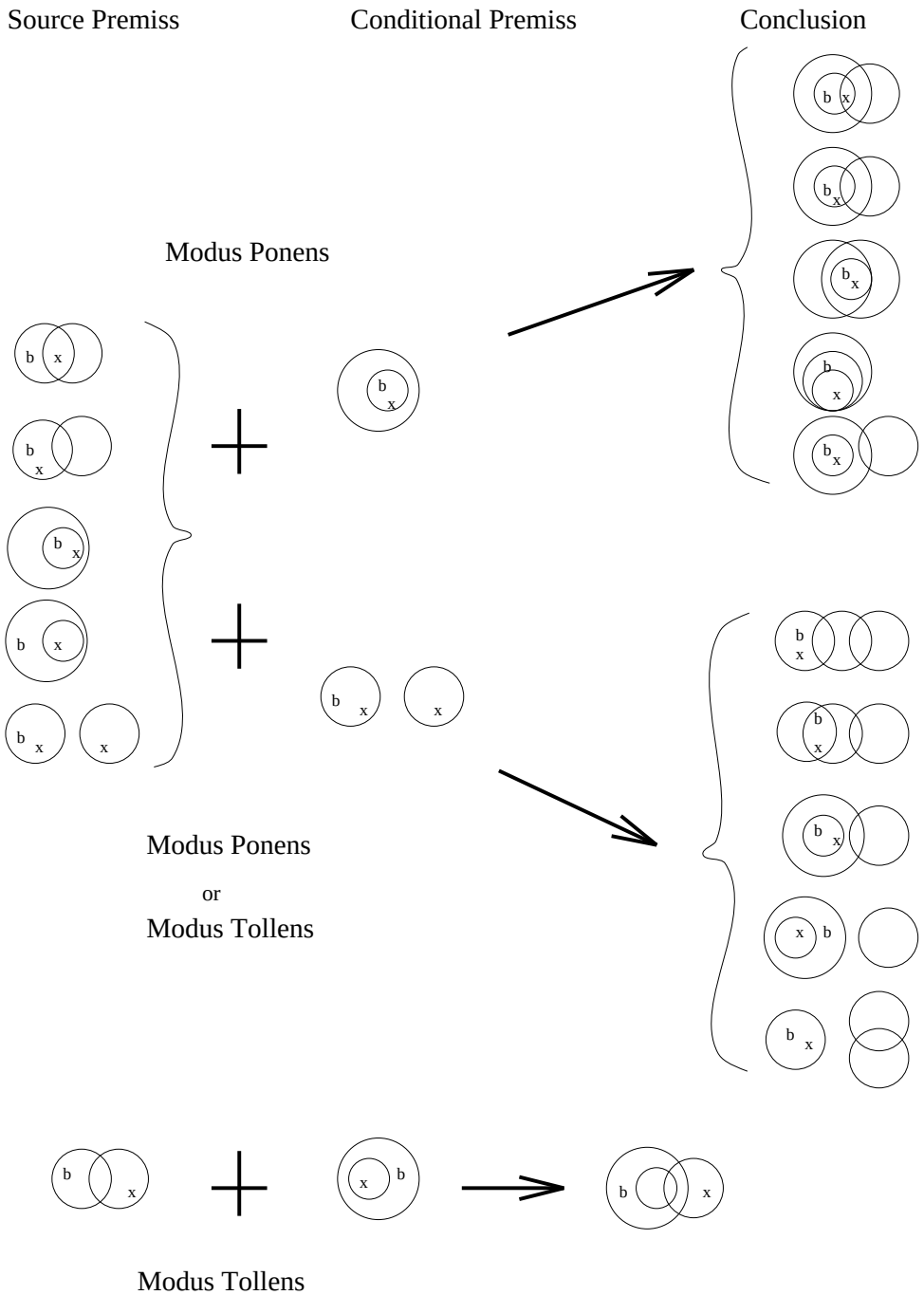

Figure 6: Graphical correspondences to sentential rules: the 5 diagrams in which b contains an $\times$ are registered with either of the two universals with B as a minimal region in the analogue of *modus ponens*. The one diagram in which b contains no $\times$ is registered with the one universal with b as a non-minimal region in the analogue of *modus tollens*. The two rules yield the 10 diagram-types with valid conclusions in Figure 12

are not C. Considered graphically, the critical reasoning operation is to register the existential b-circle containing a $\times$ outside its intersecting c-circle, with the universal b-circle wholly inside the a-circle. To realise that the existential $\times$-region will not be cut by the a-circle we do not have to further consider the three possible arrangements of c relative to a. The constraint that the existential $\times$ remains outside c is sufficient to validate the inference. When Euler's algorithm is conceived this way, the graphical and sentential algorithms coincide completely.

This presentation emphasises several generalisations about syllogistic reasoning. At their core, syllogisms are conditional inferences. In these inferences one premiss supplies the conditional and the other the source of the atomic input to the conditional rule. We return to these distinctions in analysing the empirical data of human reasoning.

Summary In these logical investigations, we have shown that the syllogism has the property of case-identifiability. We have described a general class of algorithms we have called individual identification algorithms, and described a model-theoretic algorithm (implemented in `prolog`), and graphical and sentential implementations of this family of algorithms. Case identifiability plays a central role in determining the applicability of this abstract algorithm. The general class of algorithms, and especially the original statement of the graphical version, assumes complete determination of both minimal and maximal models for premisses. The sentential algorithm applies a more focussed strategy predicting the existence and character of critical individuals. But it is not difficult to recast the graphical algorithm so that it only computes the maximal model. Once recast, it is fully isomorphic with the EC algorithm.

We now turn to the implications of these investigations for the empirical study of human syllogistic reasoning.

3 Using the Framework to Analyse Performance

We have so far been concerned with the logic of syllogistic reasoning rather than any performance model of actual human reasoning. We now ask how our novel analyses of the logic can contribute to the modelling of human mental processes.

There are two chief kinds of data which previous theories have modelled—errors in reasoning, and *figural effects* or choices between alternative structures for conclusions. Johnson-Laird & Steedman (1978) made distinctive contributions in both areas. They showed that their distinction between one, two and three model problems was a good predictor of the number of subjects making errors (failing to find valid conclusions or finding invalid conclusions), and their novel choice of task (drawing a conclusion rather than accepting or rejecting an experimenter's conclusion) meant that they presented the first systematic data on the favoured structure of conclusions, making it possible to identify the Figural Effect. Both threads have been expanded and revised in subsequent work on what has come to be known as Mental Models theory (Johnson-Laird (1983), Johnson-Laird & Bara (1984), Byrne & Johnson-Laird (1991)).

Both these kinds of data should be reconsidered in the light of the analysis presented here of relations between different implementations of the IIA. Stenning & Oberlander 1995 showed that Johnson-Laird's Mental Models method is strongly equivalent to the graphical algorithm discussed in Section 2 above. This result means that it is possible to translate the operations of Mental Models theory directly into graphical terms. Multiple-models problems are just those which have critical regions which are not bisected in some non-maximal registration,

though they are bisected in the maximal registration; errors in reasoning are thus explicable as failure to register diagrams maximally. Our present mapping of the sentential algorithm onto the graphical one reveals correspondences between mental models explanations of error and those of earlier theories such as the Chapmans' (1959) 'conversion' theories. Illictly converting $A \rightarrow B$ into $B \rightarrow A$ corresponds in mental models theory terms to failing to consider the full range of models for a pair of premisses (the dominant mode of error explanation in MM theory). Lee (1987) showed that there is a close correspondence between the three model problems and the ones where conversion theory predicts errors. The exact details will depend on the variety of conversion theory and mental models theory adopted.

Figural effects in human syllogistic reasoning are response biases in the order of end terms in conclusions, which are correlated with the figure of the premisses (Johnson-Laird & Steedman 1978). The effect can be summarised roughly as the tendency for end terms in conclusions to retain their grammatical status from the premiss in which they occur. Thus when the figure is ab/bc, ac conclusions are commonest, ba/cb problems give rise to ca conclusions, and there is less overall bias when the figure is ab/cb or ba/bc. These are highly reliable effects, which have been shown to occur in both valid conclusions (where the conclusion is convertible) and invalid conclusions.

Mental models theory explains these effects in terms of the first-in-first-out (FIFO) architecture of working memory (Baddeley 1986). Terms are held in a pipeline memory and can only be reordered by extensive reversal operations. This explanation appeals to a very particular implementational detail, though without any empirical test through memory experimentation. What evidence there is about FIFO memory deployment in syllogistic reasoning (Gilhooly *et al* 1993) fails to show expected suppression effects, though this may be due to methodological problems of incomplete suppression.

But such appeals to details of memory implementation are premature when there is so wide a range of types of implementation which are not distinguishable on the basis of input/output mappings. Natural implementations of the IIA in both graphical and sentential modes generate ordered term structures without making any highly specific assumptions about the substrate of term-order memory. It is clear from the published data (e.g. Johnson-Laird & Steedman 1978, Johnson-Laird & Bara 1984) that the grammatical organisation of conclusions interacts powerfully with the quantifiers as well as with figure. What is needed is an explanation of the how the grammatical structuring-effects result from the processes of reasoning rather than a characterisation of the substrate which preserves order information, especially when it is clear that several substrates may equally well achieve the same effect.

In this paper we pursue both threads of investigation, but concentrate our efforts on modelling term order effects with a view to explanation of the figural effect. We make this choice because our identification of the IIA suggests a way of generating richer empirical data. The equivalences we have noted between apparently distinct algorithms makes us sceptical about differentiating these implementations on weak behavioural observations, especially on the grounds of input/output mappings between premisses and conclusions. The empirical base of measurements of human performance in this area is severely impoverished. Further, our definition of an abstract algorithm and demonstration of correspondences between that algorithm and many extant models of human performance raises an important empirical question prior to questions of which implementation of the IIA might be employed by subjects. That is the question, do human subjects use the IIA at all? Or have previous theories been mistaken in covertly assuming that this is how people reason? Figural Effect data from a novel task can provide strong evidence that indeed some implementation (or implementations) of

the IIA is spontaneously adopted by human reasoners in order to solve this task.

The case-identifiability property of syllogisms, and the IIA constructed on its semantic basis, suggest both an explanation of ordering effects at an appropriate level of abstraction, and a method of gaining much richer data about mental processes. The IIA is a method for deciding whether there is any maximally specified individual which must exist in any model of the premisses, and of constructing a specification if any does exist. What distinguishes the algorithm is precisely the division of the reasoning process into a part which builds up specifications of fully determined individuals, and a part which draws conclusions by abstracting away from these specifications. The task which most directly operationalises the first half of this process is the task of giving descriptions of maximally specified individuals whose existence is entailed by the premisses. We call this the *individual task*.

This task has two main virtues for studying figural effects in reasoning. The sequence of terms in the description of such individuals is always logically immaterial—existential conjunctions of terms are indifferent to reordering of conjuncts. In the conventional task of drawing abstract conclusions, not all conclusions are validly convertible, and so sequence of terms is confounded with validity of conclusion. Additionally, the modified task yields data on all three terms of the syllogism, rather than just on end terms.

It should be obvious that this *Individuals Task*, having uniform quantificational structure in conclusions, should be better suited to evaluate claims about the relative difficulty of single- and multiple-model problems than is the *Standard Task* of drawing quantified conclusions relating the end terms. The evidence for a difference in difficulty is confounded, in the Standard Task, by the quantificational structure of conclusions, since multiple-model problems also tend to have conclusions in a different mood from either of the premisses, while single-model problems usually do not (Ford 1994).

Finally, the Individuals Task allows us to investigate whether naive reasoners agree with Aristotle that nothing follows from two negative premisses. If subjects can find the maximal individuals entailed by the premisses which licence U-conclusions, that provides evidence that the IIA underlies their reasoning.

We first elaborate an account of figural effects in Subsection 3.1 in terms of the concept of *source premiss* and test predictions in Subsection 3.2 using the novel Individuals Task. The results of this experiment are used to guide the construction of a model of the ordering of all three terms in individual conclusions.

3.1 Individuals and Figural Effects

The IIA casts the syllogistic reasoning task as one of constructing three-term descriptions of critical individuals. In a finished diagram these descriptions correspond to $\times$-marked regions. In the sentential version of the algorithm, they are the three-conjunct statements completed by application of MP or MT followed by conjoining of the concluded clause.

In the graphical version of the algorithm, the source premiss is identified as the premiss contributing the $\times$ which persists into the final diagram. In the sentential algorithm, each of the routes to the application of MP or MT identifies one of the premisses as source premiss. One logical generalisation about source premisses is particularly noteworthy—if there is an existential premiss, then *it* is the source premiss—but since universals have existential consequences in the conventional interpretation of the syllogism, these can be source premisses too. In general, the source premiss for some valid individual is any premiss which makes an existential assertion, from which the existence of that individual can be inferred by using the

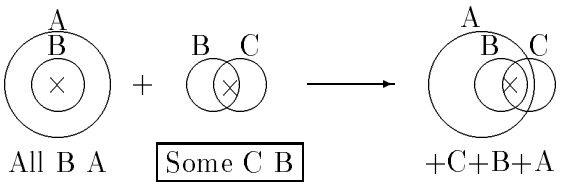

Figure 7: Registration diagram for a problem with source Premiss 2

other (*conditional*) premiss as a rule of inference. For each valid individual for each syllogism, that individual may specify a unique source (e.g. Figure 7), or fail to because more than one derivation of that individual from those premisses is possible.

The source premiss assignment of problems is empirically interesting because it is distributed unevenly among the four figures, in a way that appears to parallel the figural effect. That is, there is a tendency for ab/bc conclusions to have source Premiss 1, ba/cb conclusions to have source Premiss 2, and for the symmetric figures to have equal numbers of conclusions with each source premiss. This is a reflection of the *logical* figural effect in the Standard Task, where there is an imbalance in the numbers of valid conclusions with each possible term order in each figure. In ab/bc, for example, there are more unconvertable ac valid conclusions than ca ones. The distribution of source parallels this logical figural effect, so that, usually, unconvertable ac conclusions have source premiss 1, and unconvertable ca conclusions have source premiss 2. The exceptions are the vn2 problems which have unconvertible conclusions but no unique source.

However, there are also convertible conclusions with unique source, and in these cases it appears that the source premiss predicts human reasoners' preferences quite well. In the second premiss source problem in figure 7, for example, the overwhelmingly most popular conclusion is "Some of the Cs are As", even though the converted form "Some of the As are Cs" is equally valid (Johnson-Laird 1983).

The confounding influence of unconvertible valid conclusions makes the Standard Task an unsatisfactory testbed for hypotheses about source premisses. However in the Individuals Task, in which all conclusions are convertible, we expect source to give rise to term order phenomena comparable to the Standard Task figural effect, both for problems whose Standard Task conclusions are convertible, and those which are not. In particular, *we expect that the end term from the source premiss will tend to precede the other end term in conclusions*, but given that the IDs contain the middle term b as well as the end terms, we expect the terms from the source premiss to constitute the head of the description, possibly in their premiss order, followed by the remaining end term. The uneven distribution of problem types should then give rise to an end-term order pattern similar to the overall figural effect. Thus we might be able to explain the figural effects as the consequence of a strategy of beginning the construction of the ID from the source premiss.

3.2 The experiment

The experiment collected reasoning data on performance in the novel Individuals Task, and the Standard Task—drawing quantified conclusions relating the end terms—for comparison with previous published results.

In the Individuals Task condition, subjects were instructed to determine, for any premiss pair, whether there was any type of maximal individual which *must exist*, given the truth of

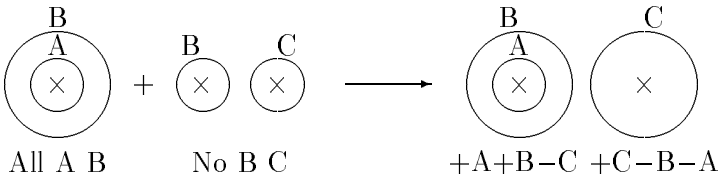

Figure 8: Registration diagram for a problem with two valid individuals, each with a different source.

the premisses, and if so to describe it. Responses took the form of lists of features, which specify for each term in the problem, whether the individual does or does not have the property denoted by the term.

In the Standard Task condition, subjects were instructed to draw quantified conclusions relating the end terms. The data from the Standard Task are used for a problem-by-problem comparison with the Individuals Task data, to establish the relation between the performance profiles of the two tasks. The aim is to assess the degree of psychological, as opposed to logical similarity between the tasks. One basis for comparison is the difference in difficulty between single- and multiple-model problems; this can serve as a benchmark for evaluating performance on U-conclusion problems, against performance on conventionally valid multiple-model problems.

The remaining analyses examine term ordering phenomena in the Individuals Task condition. We seek to examine the novel data on middle term position in Individual descriptions, and to test predictions of the effects of source premisses on term order.

If a subject correctly draws a valid conclusion to a problem, we can determine its source, providing that a unique source is defined for that conclusion to that problem. We predict an association between source premiss and end-term order, such that when the source is Premiss 1, a will precede c in the conclusion, and when the source is Premiss 2, c will precede a. When either premiss could have been the source, we make no prediction.

There is a set of problems which have two different individual conclusions, each with a different source. These have an A premiss whose subject is an end term, and an E premiss (e.g. AabEbc - Figure 8 shows the registration diagram). Each of these individuals supports a different quantified conclusion, having only one positive feature corresponding to an end term (e.g. +a+b−c supports Eac (or Oac) and +c−b−a supports Eca (or Oca)). Since we cannot state *a priori* which premiss will be source, we cannot predict an overall figural effect in these problems, 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 drawn. Specifically, conclusions whose source is the A premiss should have the end term from that premiss before the end term from the other, and conclusions whose source is the E premiss should have the opposite end term order.[4]

The dependency between source and end-term order should hold for problems in all the figures. In the ab/bc and ba/cb figures the imbalance in numbers of problems with Source 1 and 2 should give rise to an apparent effect of figure : ab/bc has more Source 1 problems, so it should show an overall ac bias, whereas ba/cb, with more Source 2 problems, should show a ca bias. There are problems in both the ab/cb and ba/bc figures which have unique source, but

[4]Such a *post hoc* identification of source premisses would not be possible for Standard Task conclusions, since Eca and Eac imply each other; however the Individuals Task conclusions are logically independent.

owing to the existence of complementary pairs of problems, with the same logic but inverted premiss order, there should be no overall bias due to the distribution of source in these figures. The size of the diagonal effect can be estimated by assuming all correct conclusions to be equally likely, then the relative numbers of conclusions with each source premiss in the diagonal figures predicts a small Figural bias in the region of ±7-10% (depending how we count conclusions to the two-individual problems).

The consequence of preservation of term order from the source premiss (or indeed any founding premiss) in Individual Descriptions is that in each figure, we predict two dominant three-term orders out of the possible six, one founded on each of the premiss orders which jointly constitute that figure. Since these always have the middle term b in first or second position, there should be few conclusions with b final.

3.2.1 Method

Design Separate groups of subjects produced either descriptions of necessary individuals, or quantified conclusions relating the end terms, for each of the 64 syllogisms presented in random order.

Subjects 22 Edinburgh University students, and 24 Durham University students, took part in the Individuals and Standard Task conditions respectively. None had any prior training in syllogistic logic.

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 on each slip of paper described a group of people at a party. They were instructed to assume that some people corresponding to each of the three terms existed.

Subjects in the Individuals Task condition were instructed to decide whether any kind of person who could be described with certainty, in terms of either positive or negative values of all three features, had to be present in the room, and to describe the individual on the slip of paper, or if there was no such individual, to write "No valid conclusion".

Subjects in the Standard Task condition were instructed to decide whether there was any statement relating the end terms of the premisses which must hold of the people in the room, and to write the statement on the slip of paper, or if there was no such statement, to write "No valid conclusion".

Subjects worked individually in quiet surroundings, and were given as much time as they needed to finish all the problems.

3.2.2 Results

Reasoning accuracy: comparison between tasks In order to investigate the relation between performance on the Individuals Task and performance on the Standard Task, two scattergrams were constructed, showing data from problems without Valid Conclusions (Figure 9) and with Valid Conclusions (Figure 10) respectively. Each point represents a single problem, and its X and Y values represent the percentage of subjects who responded correctly to it on each task. Moreover, each problem is labelled with the name of its diagram type (see Figure 12). Since the U-conclusion problems have valid conclusions in the Individuals Task but not in the Standard Task, results for these problems on the two tasks are not directly

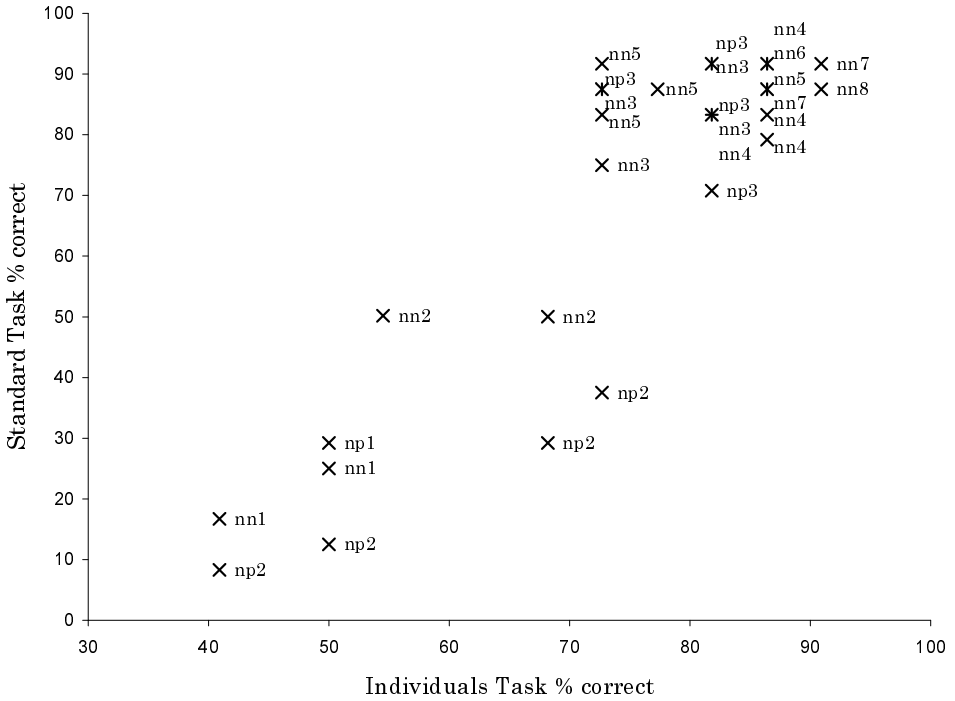

Figure 9: Scatter plot showing scores on NVC problems in both tasks.

comparable, so they have been omitted from the main analysis. However, they have been plotted on Figure 10 using a different marker type for convenience.

Figure 9 shows the data for problems without valid conclusions. Although scores on the Individuals Task are higher than those on the Standard Task, problem difficulty is highly correlated between the two tasks ($r = 0.88$, 26 $d.f.$ $p < 0.0005$). By inspection, the problems fall into two groups, one high-scoring and one low-scoring; moreover, no diagram type represented in either of the groups occurs in the other group[5]. The high-scoring group is comprised of problems from diagram types np3, nn3, nn4, nn5, nn6, nn7 and nn8, while the low-scoring group comprises diagram types np1, np2, nn1 and nn2. This pattern of results is consistent with the hypothesis that subjects use the Aristotelian generalisations, that no problem with two particular premisses has a valid conclusion, and that no problem with two negative premisses has a valid conclusion. All and only the high-scoring problems can be decided in this way, so the remaining problems would be expected to cause more difficulty, as observed. From Figure 10 it can be verified that the Standard Task scores for U-conclusion problems, where the correct response is No Valid Conclusion, fall in the same high-scoring region as would be expected on this hypothesis. Note, however, that the low-scoring problems are also those which have an *All* premiss.

Figure 10 is a plot of performance on both tasks for problems with valid conclusions. For convenience, the U-conclusion problems are represented with circles appearing in the top left region. The valid conclusion results are less highly correlated than the NVC results ($r = 0.56$, 24 $d.f.$ $p < 0.005$); this is due to a substantial subset of problems in the bottom right region with high Individuals Task scores but very low Standard Task scores. As in

[5]Diagram types abstract over the labelling of the end-term circles.

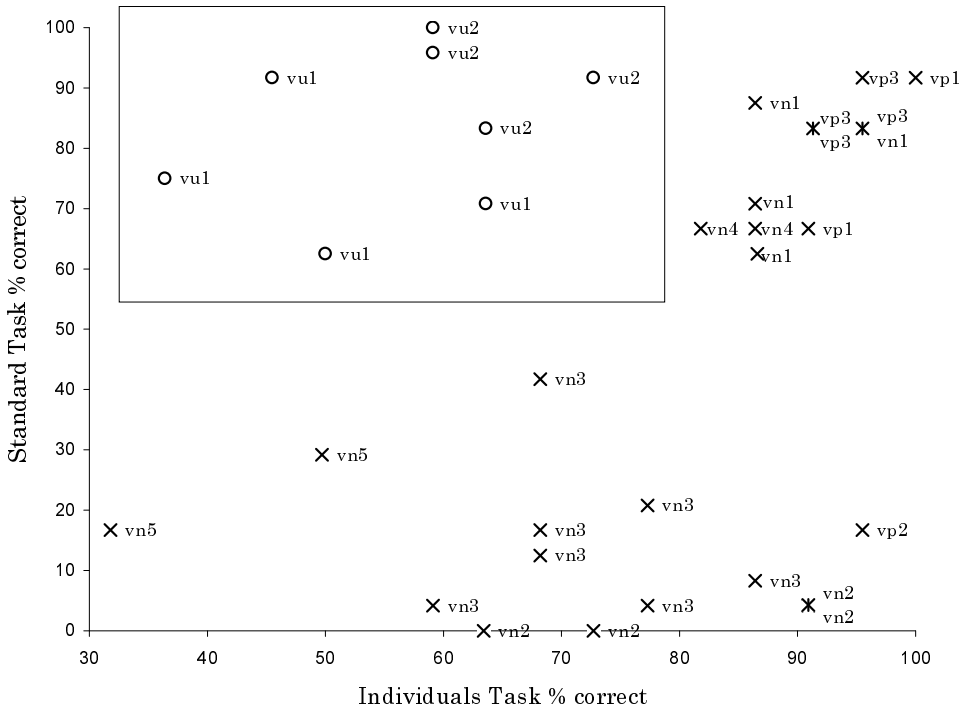

Figure 10: Scatter plot showing scores on VC problems in both tasks. Also shown are the U-conclusion problems, marked with circles in the box at top left - since these have valid conclusions in only the Individuals Task, their Standard Task scores are not comparable with the rest of the Standard Task scores on this chart.

Figure 9 there is a group of problems with high scores on both tasks: these are instances of diagram types vp1, vp3, vn1 and vn4, and as before no other instances of these diagram types appear anywhere else on the chart. All but one of the single-model diagram types occur in this group. In bottom left region there are a few problems with low scores on both tasks: both the vn5 problems occur here, and two of the eight vn3 problems. However, most of the vn3 problems occur, along with the vn2 and vp2 problems, in the region at the bottom right.

Since vp2 is a single-model problem, Mental Models theory would predict it to have a high level of Standard Task performance similar to the other single-model problems, rather than being as difficult as the multiple-models problems vn2 and vn3 as observed. Ford (1994) notes that vp2 is reliably difficult in other Standard Task experiments, and observes that it is the only single-model problem whose valid conclusion is in a different mood from either of the premisses, a property it shares with vn2 and vn3 problems. Choosing a suitable quantifier for the conclusion is a source of difficulty which occurs in the Standard task but not in the Individuals task, so it is reasonable to suppose that this different-mood hypothesis accounts for the observed difference in difficulty between the tasks for these problems. However, since most of the Standard Task multiple-models problems are vn2 and vn3 problems, the different-mood hypothesis undermines most of the evidence for a difference in difficulty between single-model and multiple-models problems.

In terms of the Individuals task alone, there is overlap between single-model and multiple-model performance; some vn3 and vn2 problems show similar high levels of performance to

	Error conclusion		*Error NVC*	
	Inds Task	Std Task	Inds Task	Std Task
single-model	2.48%	5.47%	5.62%	16.51%
multiple-models	10.62%	29.32%	17.91%	58.00%
U-multiple-models	11.98%		29.55%	

Table 1: Mean Percentages of error conclusions and error NVC responses in both the Individuals Task (N=22) and Standard Task (N=24), for single-model, multiple-models and U-multiple-models problems (Individuals Task only).

single-model problems. However the variation in multiple-model problem scores is much greater than in single-model scores. From the graph it is clear that while generating valid U-conclusions is fairly difficult, there are other multiple-model problems with comparable levels of difficulty. One of the vn5 problems shows lower scores than any U-conclusion problem, and the spread of U-conclusion problems overlaps to a considerable extent the distribution of vn2 and vn3 problems. It is not unreasonable, therefore, to consider the U-conclusion problems as multiple-models problems.

We can use a mixed-model Analysis of Variance to investigate the effect of number of models (repeated measures) on percentage error rates, for both Individuals Task and Standard Task subjects. Two ANOVAs were conducted, using percentage of error conclusion responses and error NVC responses respectively as the dependent variables. Table 1 shows for both Tasks, how mean percentages of error conclusions and error NVC responses vary between single-model, multiple-models and U-multiple-models problems.

Percentages of error conclusions are higher in multiple-models problems than in single-model problems, for both tasks ($Overall: \ F(1,44) = 34.84, \ p < 0.0001$). There is also a main effect of Task, such that there are more error conclusions in the Standard Task than in the Individuals Task ($F(1,44) = 11.07, \ p < 0.0018$), and there is a significant interaction between number of models and task ($F(1,44) = 8.40, \ p < 0.0058$). Although the difference between single-model and multiple-models problems is smaller in the Individuals Task, it is nonetheless significant ($t(21) = 2.67, \ p < 0.01$); however there is no significant difference in percentages of Error Individuals between multiple-models and U-multiple-models problems ($t(21) = 0.35, \ n.s.$).

Similarly for Error NVC responses, there are main effects of number of models ($F(1,44) = 95.22, \ p < 0.0001$), task ($F(1,44) = 27.23, \ p < 0.0001$) and a task by number of models interaction ($F(1,44) = 28.08, \ p < 0.0001$). Again, despite the interaction, there are significantly more error NVC responses to multiple-models problems than to single-model ones in the Individuals Task ($t(21) = 4.14, \ p < 0.0005$), but also there are significantly more error NVC responses to U-multiple-models problems than to conventional multiple-models problems ($t(21) = 2.30, \ p < 0.025$).

So the Individuals Task shows an effect of the number of models which is similar to, but less extreme than, the effect familiar from the literature on the Standard Task: multiple-models problems cause more errors of both types than do single-model problems. The U-multiple-models problems, which only occur in the Individuals Task, differ from conventional multiple-models problems only in having a higher rate of Error NVC responses. We can conclude that for the Individuals Task data, although subjects apparently do sometimes use the Aristotelian generalisation that nothing follows from two negative premisses, this is by

			figure				
			ab/bc	ba/cb	ab/cb	ba/bc	Overall
Individuals Task	valid	ac %	65.9	49.2	45.5	57.1	55.8
		N	138	122	88	217	565
	invalid	ac %	73.0	39.7	64.9	72.2	61.2
		N	63	63	57	36	219
	Overall	ac %	68.2	45.9	53.1	59.3	57.3
		N	201	185	145	253	784
Standard Task	valid	ac %	88.2	8.8	54.2	51.2	50.4
		N	68	68	48	86	270
	invalid	ac %	90.0	15.5	53.7	67.7	56.7
		N	90	84	82	65	321
	Overall	ac %	89.2	12.5	53.8	58.3	53.8
		N	158	152	130	151	591

Table 2: The Figural Effect in both tasks: percentages of valid and invalid conclusions with ac term order in each figure.

no means universal, and many do successfully draw the valid U-conclusions. In other words, the different task instructions do succeed in switching many subjects' behaviour.

Conclusion Term Order: a comparison of tasks The most straightforward way to compare term order in the Individuals Task with that in the Standard Task is to initially consider only the order in which the end terms occur in individual descriptions. Tabulating the percentages of conclusions in which a precedes c in the four figures, for both valid and invalid conclusions, in each Task, we derive Table 2, which shows the Individuals Task counterpart of the figural effect. Taking account of a slight overall bias towards ac conclusions, it is clear that there is an overall Individuals Task figural effect ($\chi^2(3) = 20.88$, $p < .0001$) similar to the type reliably observed in Standard Task experiments. It is less pronounced for valid than for invalid conclusions, but is significant in both cases [Valid: ($\chi^2(3) = 11.90$, $p < .0077$), Invalid: ($\chi^2(3) = 18.16$, $p < .0004$)]. For valid conclusions in the diagonal figures, after taking account of an overall ac bias of 6-8%, the size of the effect (±8-10%) is similar to that predicted by the distribution of source in these figures. The Standard Task figural effect is more extreme than this. While the figural effect in valid conclusions arises in part from the imbalance in numbers of valid conclusions with each term order in the diagonal figures, the size of the effect in invalid conclusions is also larger.

Table 3 shows how end-term order varies with source premiss. The Table shows ac order for problems with unique source, problems in which both premisses are source, and invalid conclusions. End-term order is strongly associated with source ($\chi^2(2) = 118.1$, $p < 0.0001$), 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, as predicted.

Table 4 shows the relation between end-term order and source premiss for all of the problems which establish two different individuals (e.g. Figure 8). On the basis of the conclusion, 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

	source			
	Premiss 1	Premiss 2	Both	*Overall*
ac %	77.9	28.0	65.7	55.8
N	213	218	134	565

Table 3: The Source Effect in the Individuals Task: Percentages of valid conclusions with ac order in first- and second-premiss unique source problems, and non-unique source problems.

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

Table 4: Association between end-term order and source premiss for valid conclusions to problems which establish two individuals. End-term order is AE if the end term from the A premiss precedes that from the E premiss.

the Table shows, there are more conclusions 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 ($Yates'\ \chi^2(1) = 46.24,\ p < .0001$).

Table 5 shows how well source predicts end-term order in each figure. In the first three figures, source predicts end-term order very well, with more than 90% of problems with unique first-premiss source having ac conclusions; second-premiss source predicts ca conclusions in these figures slightly less well, but much better than chance. Possibly this difference is attributable to the small overall preference for ac conclusions. More seriously however, source predicts end-term order in ba/bc at chance levels.

Our final predictions concern the novel data on the position of the b term in IDs. The hypothesis that Individual Descriptions are founded on one of the premisses, followed by the end term of the other premiss, predicts two orders in each figure. Table 6 shows the percentages of responses with each possible term order in each figure, with the cells corresponding to each predicted order emboldened. It is clear that in each case as predicted, the emboldened cells do contain the highest proportions of responses, by a considerable margin. As expected, there are few responses with b final, and responses with b initial are slightly more frequent than those with b medial.

The preceding analyses show that for most of the relevant data, source premiss is a good predictor of end-term order. The exception cases are in figure ba/bc, where the association

	figure			
	ab/bc	ba/cb	ab/cb	ba/bc
source 1 ac%	93.8	92.6	90.0	50.7
source 2 ca%	80.0	78.1	91.7	47.8

Table 5: Prediction of Individuals Task end-term order by source in each figure

	order						
figure	bac	bca	abc	cba	acb	cab	*N*
ab/bc	7.5	**23.9**	**59.2**	3.5	1.5	4.5	201
ba/cb	**37.3**	6.5	4.3	**44.3**	4.3	3.2	185
ab/cb	13.1	4.1	**36.6**	**35.2**	3.4	7.6	145
ba/bc	**52.6**	**35.6**	5.5	3.2	1.2	2.0	253
Overall %	**30.1**	**19.9**	**24.7**	**18.9**	2.4	4.0	784

Table 6: Percentages of Individuals Task conclusions with each possible term order in each figure. Cells predicted to be the largest in each row are emboldened.

is at the level of chance. Closer examination of individual problems in this figure reveals that when the conditional premiss has an A quantifier, slightly more individuals appear to be founded on the terms from the A premiss, followed by the source end term, than are founded on the source premiss in the manner predicted. Table 10 in the Appendix shows an example set of problems, two in the diagonal figures and two in ba/bc. The first two, on the left, are paradigmatic examples of the source effect (and the figural effect), while the two on the right appear to invert the predictions of the source effect, for both the Individuals and Standard Task data.

We call this tendency to found conclusions on the terms of the A premiss the A-Effect. Wider examination reveals that problems with conditional A premisses in other figures usually have at least a few conclusions apparently founded on the A premiss. Furthermore, it can be detected in other published datasets, such as those in Johnson-Laird & Bara (1984), and Johnson-Laird & Steedman (1978), as one or both of: a main A-Effect irrespective of figure, and an interaction with figure such that the association of term order with A premisses is greatest in ba/bc and least in ab/cb (Yule 1995). In view of these considerations, we have included the quantifier of the conditional premiss as a variable in a model of three-term order effects presented in the next section.

A loglinear model of three-term order effects The following analysis uses only valid conclusions with unique source premisses (diagram types vp1, vp3, vn1, vn3, vn4, vn5 and vu2). The multiway frequency tabulation uses the independent variables **source** (first/second), **srcgram** (source grammar: endsubj/midsubj), **condgram** (conditional premiss grammar: endsubj/midsubj) and **condq** (the quantifier of the conditional premiss: all/none) against the dependent variable **order** (order of all three terms in the individual: abc, bac, cba, bca). Owing to sparsity of data, the 9 responses with term b final are omitted.

The main overall hypothesis is that individuals will be constructed by taking the terms from one (founding) premiss in order, and adding the remaining end term from the other premiss last. We predict that there will be a tendency for the founding premiss to be the **source** premiss. For the first part, we predict similar main effects of **srcgram** and **condgram**, such that midsubj biases for conclusions with the middle term first, while endsubj biases for conclusions with the middle term second. For the second part, when the **source** is the first premiss, we expect biases towards conclusions in which a precedes c, and when the **source** is the second premiss, there should be biases towards conclusions in which c precedes a. The presence of the variable **condq** allows us to take account of the A-Effect, by allowing the effect of **source** to interact with the quantifier of the conditional premiss.

		order			
		abc	*bac*	*cba*	*bca*
		0.242	0.018	-0.474	0.214
	source				
	first	**0.751**	**0.105**	-0.698	-0.157
	second	-0.751	-0.105	**0.698**	**0.157**
	srcgram				
	endsubj	**0.722**	-1.020	**1.130**	-0.832
	midsubj	-0.722	**1.020**	-1.130	**0.832**
	condgram				
	all	**0.263**	-0.399	**0.565**	-0.429
	none	-0.263	**0.399**	-0.565	**0.429**
	condq				
	all	0.021	-0.042	0.173	-0.153
	none	-0.021	0.042	-0.173	0.153
source	**srcgram**				
first	endsubj	**1.333**	-0.576	-0.961	0.204
	midsubj	-1.333	**0.576**	0.961	-0.204
second	endsubj	-1.333	0.576	**0.961**	-0.204
	midsubj	1.333	-0.576	-0.961	**0.204**
source	**condgram**				
first	endsubj	-0.511	0.489	**0.399**	-0.378
	midsubj	0.511	-0.489	-0.399	**0.378**
second	endsubj	**0.511**	-0.489	-0.399	0.378
	midsubj	-0.511	**0.489**	0.399	-0.378
condq	**source**				
all	first	-0.558	-0.562	0.717	0.403
	second	0.558	0.562	-0.717	-0.403
none	first	**0.558**	**0.562**	-0.717	-0.403
	second	-0.558	-0.562	**0.717**	**0.403**

Table 7: Loglinear parameters for each effect in Model 1 ($LR\ \chi^2(21) = 30.17,\ p < 0.0886$). Boldface indicates parameters predicted to be positive.

Table 7 shows the loglinear parameters for all effects included in the simplest model with no significant lack of fit. This includes the interactions **srcgram** × **source** × **order**, **condgram** × **source** × **order** and **condq** × **source** × **order**, as well as their component lower-order effects.

From the table, all the predicted main effects are in the expected directions. Parameters for the **srcgram** × **order** effect are considerably larger than those for the **condgram** × **order** effect. The presence of the **source** × grammar interactions (**srcgram** × **source** × **order** and **condgram** × **source** × **order**) shows that the grammar and **source** effects cannot be considered independent. Each has three degrees of freedom, permitting 4 predictions each. These are shown in boldface, and all predicted parameters turn out positive as expected.

We include the parameters for the main effect of **condq** for completeness, but these are small and have no obvious interpretation in terms of our hypothesis. From the table of parameters for the **condq** × **source** × **order** effect, we can see that the predicted **source** effect is greater when the conditional quantifier is 'none', and is less when the conditional quantifier is 'all', as expected on the basis of the results of the previous section.

4 Errors and the A-Effect

The three-term model shows that as well as the predicted founding of conclusions on the source premiss, there is also a competing tendency to found conclusions on A premisses[6]. We note that one way of discriminating the difficult NVC problems is in terms of whether they have A premisses. It is possible that a single mechanism could explain both these phenomena, for example a solution strategy which is unsound in general but yields valid conclusions in many cases. We next propose and test some hypotheses about what these unsound strategies might be.

While the unsound heuristics most frequently discussed in the literature, such as the Atmosphere Effect (Woodworth & Sells 1935; Revlis 1975) and Matching (Wetherick 1989; Wetherick & Gilhooly 1990) cannot be easily applied to the case of the Individuals Task, since they are mainly addressed to the problem of quantifier selection, the Conversion Hypothesis (Chapman & Chapman 1959; Revlis 1975; Newstead 1989, 1990) is straightforwardly applicable. The Individual Identification Algorithm can be adapted to make conversion errors simply by changing the characteristic representations of premisses. In graphical terms, this is achieved by using a single circle to represent both terms of the A premiss—this strategy would generate correct answers to problems with A premisses and valid conclusions, but would yield extra invalid conclusions in np1, np2, nn1 and nn2. The IIA Conversion Hypothesis cannot, in itself, account for the A-effect on term order, because processing would proceed in the usual, source-first fashion; even in a natural deduction system, since conversion changes the order of the terms of the *conditional* premiss, it is not obvious that this would lead to any term order effects different from those predicted by the source-founding hypothesis.

Another possible explanation derives from previous work on the Individuals Task dataset. Yule (1991) demonstrated that the majority of error individuals could be construed as simple unifications of the critical individuals from the two premisses. This tendency was most marked in problems with at least one A premiss. Thus with premisses "All As are Bs, Some Bs are Cs", the A premiss provides the critical individual +A+B, which is simply unified with the critical individual from the I premiss, if possible (i.e. provided the middle term features match

[6]See Yule (1996) for evidence that the A-effect also occurs in Standard Task datasets.

Problem Type	Error Inds	Conversion Attributable	A founded	Source founded
np1 (1)	11	9	(7)	
np2 (4)	36	36	17	4
nn1 (2)	22	21	19	0
nn2 (2)	15	4	0	3

Table 8: Errors attributable to conversion.

in polarity), to give the erroneous conclusion +A+B+C. We can call this operation *Minimal Linking*, and obviously it will often return the valid conclusion. For example, provided the middle terms match, and the conditional is an A premiss, Minimal Linking will usually return the valid individual if one exists. In such cases, if no valid conclusion exists, the Minimal Linking individual will be the same as that predicted by conversion.

The Minimal Linking Hypothesis is distinguishable from the Conversion Hypothesis because it does not predict errors in nn2 problems (e.g. All B A, Some C not B), since the middle terms do not match in polarity, but in the other NVC problem types it predicts the same error individuals as the Conversion Hypothesis does. Furthermore, a Minimal Linking model can naturally accommodate the term-order effects, since the strategy does not depend on a source/conditional distinction, so all else being equal, conclusions founded on either of the premisses would be expected to be equally likely.

Unfortunately, nn2 differs from the rest of the problem types in which conversion theory predicts errors; we can think of the two categories as *Modus Ponens* problems (np1, np2 and nn1) and *Modus Tollens* problems (nn2). In valid *Modus Ponens* problems, the terms in the conclusion have the same polarity as their occurrences in the A premiss, whereas in *Modus Tollens* problems, the terms in the conclusion are the negations of their occurrences in the A premiss. Minimal Linking concludes that there is no valid conclusion for *Modus Tollens* problems on the grounds that the polarity of the middle term is different in the two premisses, so it doesn't find the conversion conclusion in nn2.

Valid *Modus Tollens* problems are generally much harder for human subjects than *Modus Ponens* problems; the only valid cases are the vn5 problems and the E-source conclusions to the vn1 problems, which are among the hardest conclusions for human subjects to find (see Figure 10). So the invalid conversion conclusions in nn2 might be expected to be "harder" too. Subjects who think conversion of A premisses is legitimate are hardly likely to appreciate the validity of *Modus Tollens*.

With this in mind, we can look at the relevant data. Table 8 shows, for each of the problems concerned, the total number of error individuals, numbers of error individuals which can be attributed to conversion (and therefore to Minimal Linking where appropriate), and finally the numbers of conversion-attributable individuals which are either A-founded or source-founded (in the usual sense that the terms from the appropriate premiss occur in order before the remaining end term).

For the first three problem types, almost all error individuals (96%) are attributable to conversion, whereas for the *Modus Tollens* problem type nn2, only 27% of error individuals are attributable to conversion. In the case of problem type np1, A-founding is indistinguishable from source-founding, but for problem types np2 and nn1, 63% of conversion-attributable problems are A-founded, whereas only 7% are source-founded. By contrast, no nn2 conversion-

attributable conclusions are A-founded, and 75% are source-founded.

Despite the large difference between nn2 and the rest of the problem types in terms of the number of error individuals which are attributable to conversion, as we have seen the *Modus Ponens/Modus Tollens* distinction is a possible confound. However the fact that a majority of conversion-attributable conclusions are A-founded is incompatible with the Conversion Hypothesis, while it is compatible with the Minimal Linking Hypothesis, so on balance it would appear that the Minimal Linking hypothesis is preferable. Supposing that there is some tendency among the subject population to use Minimal Linking can then explain why the *Modus Tollens* problems are so difficult, since they are insoluble by Minimal Linking.

One final observation is that for the core problem types in the diagonal figures, the invalid conclusions predicted by the minimal linking hypothesis are all Figural ones, so this hypothesis predicts an Invalid Conclusions' figural effect, as observed in this and other datasets (e.g. Johnson-Laird & Bara 1984). As a consequence, estimates of the overall A-effect in the Standard Task analyses may be inaccurate owing to the confounding of the Figural and A-Effects in these analyses.

4.1 Discussion of the experiment

The experiment establishes that subjects are about as good at specifying the critical individuals underlying U-conclusions as they are ones underlying other multiple-model syllogisms. This observation lends some weight to the view that their reasoning revealed here in solving explicitly set syllogisms is based on their intuitive grasp of relations between sets, rather than some scholastic formulation. Although they appear to be capable of using Aristotle's dictum that two negative premisses yield no valid conclusion to identify NVC conclusions in the Standard Task, they are reasonably good at drawing the U-conclusions which breach this principle in the Individuals Task. The ease with which they adopt the task of describing critical individuals, and their higher success rate than in the logically less complex conventional task, lends some support to the view that subjects at least implicitly conceive of syllogistic reasoning in terms of the IIA.

The data on the ordering of terms in subjects' descriptions of critical individuals can be simply summarised by saying that subjects build their IDs on one premiss, and add the non-middle term from the other premiss onto its tail. The complexities of the process lie in the determination of which premiss is chosen as foundation for the ID.

A number of constraints operate on subjects' choice of premiss for founding the ID. The strongest constraint is a tendency to choose the source premiss as foundation. The normative sentential algorithm (Figure 2) reveals that the source can be identified as the unique existential premiss, or the unique universal with end-term subject (possibly allowing the use of the logical symmetry of *No* premisses to achieve uniqueness).

The other constraint is an effect of the quantifier in the conditional premiss. Conditional premisses must have either *All* or *No*. If the conditional premiss has the quantifier *All*, then there is some opposing tendency to found the ID on that premiss rather than the source premiss. We have argued that this can be attributed to an alternative processing strategy, Minimal Linking, which results in characteristic error patterns similar to those predicted by Conversion of A premisses.

This latter constraint as a feature of syllogistic reasoning is a discovery of our model building process. It is characterised primarily in terms of performance deviations from outcomes predicted by our normative algorithm. Johnson-Laird & Bara (1984) discuss a similarly

unpredicted phenomenon concerning the ab/cb and ba/bc figures, writing

> "where the conclusion was in the same mood as just one of the premises, the end term of the premise tended to play the same grammatical role in the conclusion as it did in the premise itself" (p22)

In terms of our theory, the premise so identified is the source premiss, so that the source-founding theory predicts the association in ab/cb, since the source end term retains its subject status, while in ba/bc the term order is opposite to that predicted by source-founding, and similar to what we have identified as the A-Effect.

Both Lee (1987) and Ford (1994) have argued that extending the characterisation given by Johnson-Laird & Bara (1984) to the diagonal figures would generate the figural effect, and that perhaps the symmetrical and diagonal effects have a common cause. One mechanism which can explain the phenomenon in all the figures is based on substitution of conditional premiss terms for middle terms in source premisses. This generates term orders compatible with Johnson-Laird & Bara's observation. In ab/cb, substituting into the source keeps the source end term first, while in ba/bc, substituting into the source moves the source end term into last place. The substitution model however, predicts only abc and cba orders in the diagonal figures, and cab and acb orders in ab/cb; while perhaps the former are commoner than would be expected on the 'vanilla' source-founding hypothesis alone, the latter are very rare. It is possible that our results could be explained by a mixture of source-founding and substitution strategies. Ford (1994) develops a substitution-based model for 'verbal' reasoners in which both positive and negative conditional premisses are treated similarly, whereas our results suggest that this strategy is more commonly used when the conditional premiss has an *All* quantifier.

By allowing the observation of the sequencing of all three terms in critical individual identification, the results from this task reveal an extremely simple generalisation of the figural effects of syllogistic reasoning. There is a dominating tendency for the terms of a founding premiss to be preserved in the ID. The very simplicity of this idea perhaps makes it suprising that theories such as MMs which actually hypothesise three-term sequences in their underlying representations do not hypothesise the sequence observed here. They diverge because they do not predict that middle-term subjects of source premisses will typically remain in the initial position of IDs. The underlying reason is that theories such as MMs adopt a specific division of processes between representation construction and inference reading off. MM theory 'gets the b terms into the middle' because it assumes that a particular deletion operation will be the mechanism of inference. The advantages of a more abstract characterisation are that it reveals the fact that the inferences can go on in the construction of the ID, and that the b terms do not have to be 'brought together' to be deleted—b can be deleted during inference, read off from anywhere in the ID. Describing an algorithm at an abstract level reveals what is implementation detail and what is essential to the algorithm.

A final empirical observation worth mentioning is that the surface sequence of premisses has almost no effect on the reasoning process, only being evidenced in a weak residual tendency for a surplus of ac conclusions over ca ones. This lack of effect is noted in all the experiments in the literature. In the current context it shows that subjects' processes of assigning source/conditional premiss structure to syllogisms is almost completely independent of premiss presentation order.

5 General Discussion

The semantic structure of the syllogism allows processes of inference which simply construct descriptions of single individuals. Understood at a suitably abstract level, Euler's graphical method, mental models, and our sentential formulation can be seen just as different implementations of the same abstract process of constructing critical individual descriptions. Exhaustive graphical methods (such as some psychological models have supposed Euler's method to have been) are somewhat divergent from sentential methods. But the focussed methods with modalised semantics of which Stenning & Oberlander's 1995 reconstruction of Euler's method, or the equivalent mental models methods are examples, are fully isomorphic to sentential implementations. It is striking that it is just when sentential and model-based competence methods are optimised, by avoiding redundant inference through focus on the relevant case, that they come to correspond. We return to the phenomenological implications below.

Using this abstract description of syllogistic inference to model data from two tasks throws light on both the errors of reasoning subjects make, and on the processes which lead to the sequencing of terms in their conclusions. Our model of term sequencing provides a single coherent account of all the reported effects and uncovers new effects, cast entirely in terms abstract with regard to graphical or sentential representations. The source/conditional premiss categorisation drives term sequencing, modulated by particular effects due to the quantifier *all*. The observation and modelling of 3-term sequences further reveals that specific implementational claims made by, for example mental models theory, are neither born out in the data, nor required for modelling. For example, cancellation of middle terms is an operation which mental models inherits from some specific rewriting systems but is couched at a lower level of abstraction than is warranted by the data. The abstract analysis suggests that subjects appear to use something like Aristotle's meta-logical principles to reveal when there is no valid conventional conclusion in the standard task. Yet many subjects are capable of drawing valid U-conclusions from the same problems when instructed to describe critical individuals. Comparison of the two tasks strongly suggests that the problem of choosing quantifiers for conventional conclusions is a better predictor of difficulty than the number of possible models.

Describing data at the right level of abstraction is important for discovering regularities. If we are to distinguish between graphical and sentential implementations of syllogistic reasoning we need more fine-grained detail than is provided by most published experiments. Matsuno (1987) did collect gross phenomenological reports and correlated them with reasoning performance. The main finding was that the differences emerged only on problems without valid conclusions. Ford (1994) reports thinking-aloud protocols of subjects and shows more fine-grained differences in inferential performance correlated with style of report. Some of her subjects have come close (as have some of ours) to inventing the $\times$ notation while diagramming syllogisms. Stenning, Cox & Oberlander 1995 studying students learning first order logic found that there were striking individual differences in response to graphical and sentential teaching. We have no doubts that there are important individual differences, some of which are related to the graphical vs. sentential opposition which is phenomenologically so vivid.

One response to the kind of equivalence results we have reported here is to claim that they show the inadequacy of computational modelling to capture the phenomenological data of mental life. Searle's (1980) Chinese Room argument is intended to show that computation is

too abstract to explain mental experience, and the results presented here might superficially be taken as supporting his position. Our own conclusion would be quite the opposite. What our results show is that evidence such as the data of phenomenological report (which is, after all, more behaviour) is required as input to a far more sophisticated program of phenomenological data collection and computational modelling.

We conjecture that the phenomenology of the syllogistic task is driven by the *learning* process that goes on when subjects are figuring out how to exploit the constraints in what is, in its laboratory form, a newly defined task. This learning process is best conceived as one of selecting (or if preferred constructing) a representional system within which to reason. There are large families of sentential and of graphical (and no doubt other) systems available, and subjects have to navigate this space in coming to a system they can use. Subjects have fundamental knowledge about sets, and much experience with various representation systems, but they have to craft one for exploiting the particular constraints of their new task.

We believe that if the phenomenology of the syllogism is to be explained, then it will be as a computational process at the level of representation selection, not just at the level of syllogism solution. These processes are intimately intertwined in a naive subject's behaviour. But it has often been observed that by the time an 'expert' has derived a well-honed method of solution for a class of problem, little phenomenology may be left. Experts have very poor access to their processes of reasoning. In the case of the syllogism our formal equivalence between graphical and sentential processes offer an explanation about why this might be so.

Computations can be conceived at quite abstract levels, but they can also be conceived with as much implementational detail as we can justify adding. It is hard to see what program of research is going to throw light on the rich variability of human mental experience other than one that studies increasingly complicated data of that experience and models the information processing which it reflects. One great trap is to assume that our phenomenology gives us direct access to our computations and that models which are described in terms of images or sentences must be computationally distinct. We currently lack a coherent computational characterisation of the differences between graphics and sentential systems, but one critical difference lies in the differences in their attentional and information packaging mechanisms. When we have such a characterisation, we will make great strides in understanding human phenomenology.

6 Acknowledgements

Thanks to Cath Ardin for helping with design and preparation of the materials for this experiment, and to Morten Christiansen for helping to gather the data.

The support of the Economic and Social Research Council UK (ESRC) is gratefully acknowledged. The work was partly supported by the ESRC funded Human Communication Research Centre and by an ESRC studentship awarded to Peter Yule. We also wish to acknowledge partial support from Grant # GR/J586129 from the EPSRC/DTI.

7 References

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review, 85*, pps.249–277.

Baddeley, A. (1986). *Working Memory.* Oxford: Clarendon Press.

Block, N. (ed.) (1981). *Imagery.* Cambridge, Mass.: MIT Press.
Chapman, L.J & Chapman, J.P. (1959). Atmosphere Effect Re-examined. *Journal of Experimental Psychology, 38,* 220-226.
Clocksin, W.F. & Mellish, C.S. (1984). *Programming in Prolog.* Berlin: Springer-Verlag.
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.
Euler, L. (1772). Lettres a une princesse d'Allemagne. *2* sur divers sujets de physique et de philosophie. Letters 102–108.
Ford, M. (1994). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition, 54,* 1-71.
Foster, C. (1992). Algorithms, abstraction and implementation: levels of detail in cognitive science. Academic Press. London.
Gilhooly, K.J., Logie, R.H., Wetherick, N.E. & Wynn, V. (1993). Working memory and strategies in syllogistic reasoning tasks. *Memory and Cognition, 21:1,* 115-124.
Guyote, M.J. & 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. & Bara, B. (1984). Syllogistic Inference. *Cognition, 16,* 1-61.
Johnson-Laird, P.N. & Byrne, R. (1991). *Deduction.* Hove: LEA.
Johnson-Laird, P.N. & Steedman, M.J. (1978). The psychology of syllogisms. *Cognitive Psychology, 3,* 395-400.
Larkin, J. H. Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, pps. 65–99.
Lee, J. (1987). *Metalogic and the Psychology of Reasoning.* PhD Thesis, University of Edinburgh.
Leibniz, G. (1666/1966). Dissertation on the Combinatorial Art. In G.H.R Parkinson (ed. and trans.), *Logical papers: A selection.* Oxford: Clarendon Press.
Matsuno, T. (1987). Cognitive style and representational strategies in categorial syllogistsic reasoning. *Tohoku Psychologica Folia, 46.* pps. 97–102.
Newstead, S.E. (1989). Interpretational errors in syllogistic reasoning. *Journal of Memory and Language, 28,* 78-91.
Newstead, S.E. (1990). Conversion in Syllogistic Reasoning. In K.J. Gilhooly, M.T.G. Keane, R.H. Logie & G. Erdos (Eds.) *Lines of Thinking: Reflections on the Psychology of Thought, Vol. 1* Wiley: Chichester.
Pylyshyn, Z. (1973). What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imagery. *Psych Bulletin, 80* pps. 1–24.
Revlis, R. (1975). Two Models of Syllogistic Reasoning: Feature Selection and Conversion. *Journal of Verbal Learning and Verbal Behaviour, 14,* 180-195.
Searle, J. (1980). Minds, brains and programs *Brain and Behavioural Sciences, 3*, pps. 417–424. Reprinted in J. Haugeland (ed.) *Mind Design: Philosophy, Psychology, Artificial Intelligence*, pp. 282–306. Montgomery, Vt.: Bradford, 1981
Stenning, K. & Oberlander, J. (1994). Spatial inclusion as an analogy for set membership: a case study of analogy at work. In Holyoak, K. and Barnden, J. (Eds.) *Analogical Connections.*, Hillsdale, N. J.: Erlbaum. pps. 446–486.
Stenning, K. & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: logic and implementation. Cognitive Science, **19**, pps. 97–140.

Stenning, K. Cox, R. & Oberlander, J. (1995). Contrasting the cognitive effects of graphical and sentential logic teaching: reasoning, representation and individual differences. Language and Cognitive Processes, **10 (3/4)**, pps. 333–354.
Wetherick, N.E. (1989). Psychology and syllogistic reasoning. *Philosophical Psychology, 2,* 111-124.
Wetherick, N.E. & Gilhooly, K.J. (1990). Syllogistic Reasoning: Effects of Premise Order. In K.J. Gilhooly, M.T.G. Keane, R.H. Logie & G. Erdos (Eds.) *Lines of Thinking: Reflections on the Psychology of Thought, Vol. 1* Wiley: Chichester.
Woodworth, R.S. & Sells, S.B. (1935). An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology, 18,* 451-460.
Yule, P. (1991). *An experimental investigation of a new theory of syllogistic reasoning based on Euler Circles.* MSc Thesis, University of Edinburgh.
Yule, P. (1996). *Logic and Implementation in Human Reasoning: the Psychology of Syllogisms.* PhD Thesis, University of Edinburgh.

A Appendix: Logic and implementation details

A.1 Valid individual conclusions for all 64 premiss pairs.

Second Premiss	first Premiss			
	All A B	Some A B	No A B	Some A not B
All B C	+A+B+C 1	+A+B+C 1	−A+B+C	
Some B C			−A+B+C 2	
No B C	+A+B−C 1	+A+B−C 1	−A+B−C	
	−A−B+C 2			
Some B not C			−A+B−C 2	
All C B			+A−B−C 1	+A−B−C 1
			−A+B+C 2	
Some C B			−A+B+C 2	
No C B	+A+B−C 1	+A+B−C 1	−A+B−C	
	−A−B+C 2			
Some C not B	−A−B+C 2			
	All B A	Some B A	No B A	Some B not A
All B C	+A+B+C	+A+B+C 1	−A+B+C	−A+B+C 1
Some B C	+A+B+C 2		−A+B+C 2	
No B C	+A+B−C	+A+B−C 1	−A+B−C	−A+B−C 1
Some B not C	+A+B−C 2		−A+B−C 2	
All C B	+A+B+C 2		+A−B−C 1	
			−A+B+C 2	
Some C B	+A+B+C 2		−A+B+C 2	
No C B	+A+B−C	+A+B−C 1	−A+B−C	−A+B−C 1
Some C not B				

Figure 11: Necessary individuals for each premiss pair. Numerals denote source number when unique.

A.2 Case identification in prolog

This section describes an implementation, in the programming language prolog, of the Abstract Individual Identification Algorithm described in Section 2. The following code assumes standard versions of the predicates member/2, nth/3 and select/3 as defined by Clocksin & Mellish (1984).

```
characteristic(all(A,B),[[-A,-B],[-A,+B],[+A,+B]]-[[+A,+B]]).
characteristic(some(A,B),[[-A,-B],[-A,+B],[+A,-B],[+A,+B]]-[[+A,+B]]).
characteristic(none(A,B),[[-A,-B],[-A,+B],[+A,-B]]-[[+A,-B],[-A,+B]]).
characteristic(somenot(A,B),[[-A,-B],[-A,+B],[+A,-B],[+A,+B]]-[[+A,-B]]).

syll(Premiss0,Premiss1,Conc,Source) :-
        characteristic(Premiss0,Max0-Min0),
        characteristic(Premiss1,Max1-Min1),
        nth(Source,[Min0,Min1],Min),
        nth(Source,[Max1,Max0],Max),
        member(I,Min),
        unification_test(I,Max,Conc).

unification_test(I,Max,K) :-
        bagof(Unified,J^(member(J,Max),unify(I,J,Unified)),[K]).

unify([],X,X).
unify([H|T],X,[H|NT]) :-
        select(H,X,Y),
        !,
        unify(T,Y,NT).
unify([H|T],X,[H|NT]) :-
        opp(H,Opp),
        not member(Opp,X),
        unify(T,X,NT).

opp(+A,-A).
opp(-A,+A).
```

Premiss Representation In prolog, premiss types can be mapped to characteristic representations using the predicate characteristic/2, whose first argument is a premiss and whose second is a characteristic representation, represented as a Max-Min pair. Feature structures are represented as lists of features, and models as lists of feature structures. Note that each maximal model includes the type −A−B. In graphical terms this corresponds to the background of the corresponding characteristic diagram.

Case Identification The predicate syll/4 is the top level of the syllogism solution procedure. With Premiss0 and Premiss1 respectively instantiated as structures representing the two premisses, as in the first argument of the predicate characteristic/2, it retrieves two characteristic representations, one for each premiss, selects the minimal model (Min) from one and the maximal model (Max) from the other, selects a critical individual from Min, and subjects it to the unification test. If the unification test succeeds, it returns Conc, which is instantiated as a necessary individual. Standard prolog backtracking ensures that all critical individuals for both premisses are tested in turn, so the predicate finds all necessary

individuals for the premiss pair. Whenever the predicate succeeds, `Source` is incidentally instantiated as the premiss number of the source. If there is no valid conclusion the predicate fails immediately.
`unification_test/3` tests the critical individual `I` by attempting to unify it with all members of `Max`. If the `bagof` goal returns a singleton list, only one unification is possible, so that unification denotes a necessary individual.
`unification_test/3` requires a unification predicate `unify/3`, whose first and second arguments are the feature structures to be unified, which succeeds with the third argument instantiated as their unification, if this is possible, and fails otherwise. Note that in the event of a successful unification, the features from the first argument appear before any from the second in the result, so that in its use above, terms from the source premiss occur before terms from the other premiss.

A.3 Registration Diagrams

Table 12 shows the full set of topologically distinct registration diagrams for the Graphical Individual Identification Algorithm (see Section 2.3.2). These diagrams are used as the basis for the definition of a set of equivalence classes of syllogisms (*problem types*), which abstract over premiss order and the valid conversion of I and E premisses, and which are referenced in the text by the labels given to the diagrams (vp1, vp2 etc).

A.4 Sentential representations

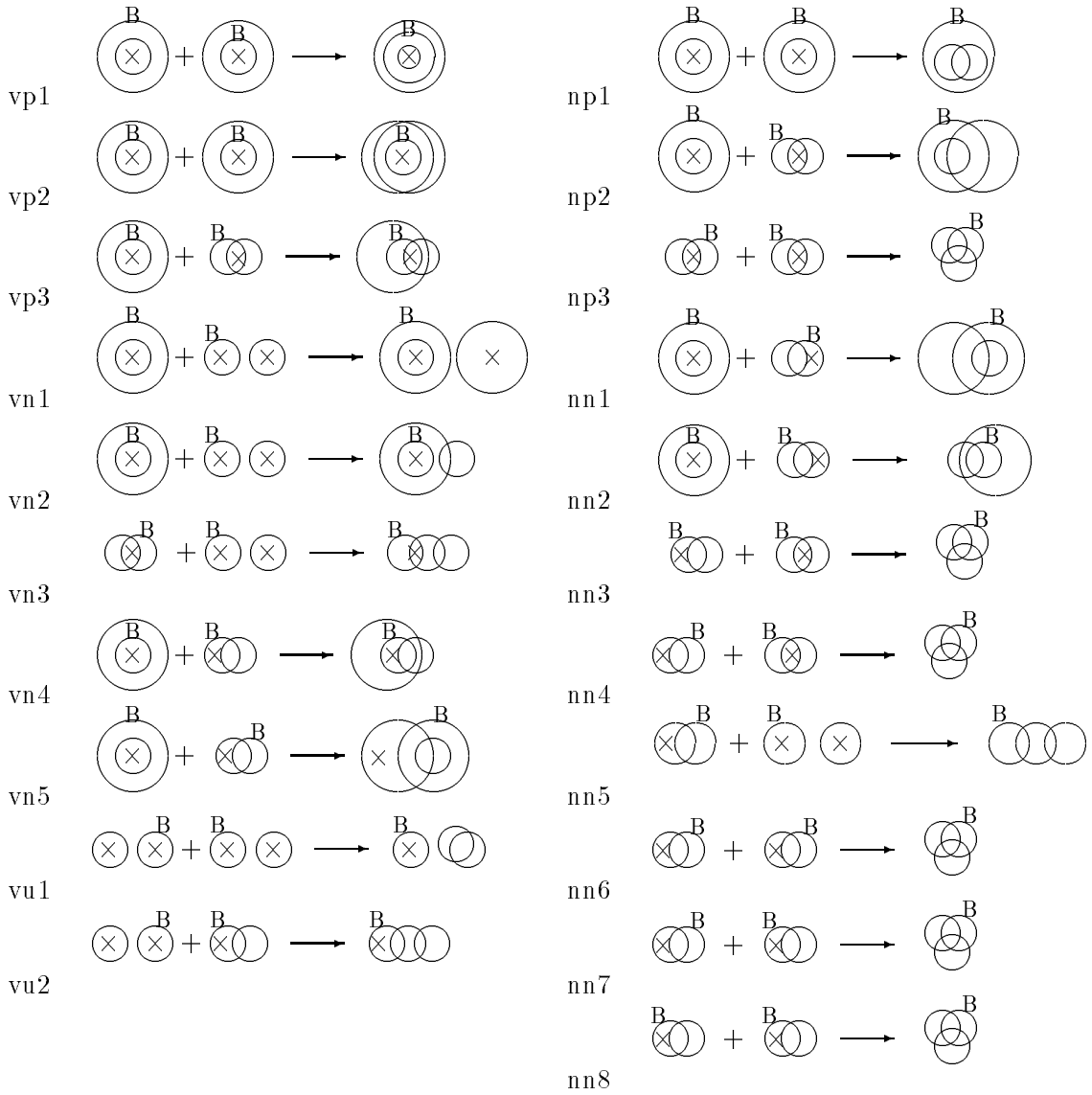

Figure 12: The full set of topologically distinct registration diagrams. Those on the left have valid individuals; those on the right do not. The unlabelled circles can represent either A or C, depending on the premisses. Crosses mark minimal regions which must be non-empty.

vp1	$\mathbf{A} \rightarrow B$	$B \rightarrow A$		
	$B \rightarrow C$	$\mathbf{C} \rightarrow B$		
vp2	$\mathbf{B} \rightarrow A$			
	$\mathbf{B} \rightarrow C$			
vp3	$B \rightarrow A$	$\mathbf{B}\&A$	$B \rightarrow A$	$A\&\mathbf{B}$
	$\mathbf{B}\&C$	$B \rightarrow C$	$C\&\mathbf{B}$	$B \rightarrow C$
vn1	$\mathbf{A} \rightarrow B$	$\mathbf{A} \rightarrow B$	$B \rightarrow \neg A$	$\mathbf{A} \rightarrow \neg B$
	$B \rightarrow \neg C$	$\mathbf{C} \rightarrow \neg B$	$\mathbf{C} \rightarrow B$	$\mathbf{C} \rightarrow B$
	(C+2MT)	MT	(A+2MT)	MT
vn2	$\mathbf{B} \rightarrow A$	$\mathbf{B} \rightarrow A$	$\mathbf{B} \rightarrow \neg A$	$A \rightarrow \neg B$
	$\mathbf{B} \rightarrow \neg C$	$C \rightarrow \neg B$	$\mathbf{B} \rightarrow C$	$\mathbf{B} \rightarrow C$
		MT		MT
vn3	$A\&\mathbf{B}$	$A\&\mathbf{B}$	$B \rightarrow \neg A$	$B \rightarrow \neg A$
	$B \rightarrow \neg C$	$C \rightarrow \neg B$	$C\&\mathbf{B}$	$\mathbf{B}\&C$
		MT		
	$A \rightarrow \neg B$	$A \rightarrow \neg B$	$\mathbf{B}\&A$	$\mathbf{B}\&A$
	$\mathbf{B}\&C$	$C\&\mathbf{B}$	$C \rightarrow \neg B$	$B \rightarrow \neg C$
		MT		
vn4	$B \rightarrow A$	$\mathbf{B}\&\neg A$		
	$\mathbf{B}\&\neg C$	$B \rightarrow C$		
vn5	$A \rightarrow B$	$A\&\neg\mathbf{B}$		
	$C\&\neg\mathbf{B}$	$C \rightarrow B$		
	MT	MT		
vu1	$A \rightarrow \neg B$	$\mathbf{B} \rightarrow \neg A$	$A \rightarrow \neg\mathbf{B}$	$\mathbf{B} \rightarrow \neg A$
	$\mathbf{B} \rightarrow \neg C$	$\mathbf{B} \rightarrow \neg C$	$C \rightarrow \neg\mathbf{B}$	$C \rightarrow \neg B$
	MT		MT	MT
vu2	$A \rightarrow \neg B$	$B \rightarrow \neg A$	$\mathbf{B}\&\neg A$	$\mathbf{B}\&\neg A$
	$\mathbf{B}\&\neg C$	$\mathbf{B}\&\neg C$	$C \rightarrow \neg B$	$B \rightarrow \neg C$
	MT		MT	

Figure 13: Sentential representations of valid syllogisms with discourse introducing property (in **bold**) and use of MT

B Appendix: Raw data tables

In the following tables, unique source premisses and conclusions founded on them are indicated in **bold**. Premisses are listed above the first underline, followed by full three-term orders for all valid Individuals Task conclusions, and finally valid Standard Task conclusions below the second underline. Numbers of NVC responses, and numbers of invalid conclusions are also included for each task.

B.1 Unique source problems

All A B All B C	All B A **All C B**
ABC (21) ACB (1)	ABC (1) **CBA (9)** ACB (1) BCA (1) BAC (8)
NVC (0); Err Ind (0)	*NVC (1); Err Ind (0)*
all A C (21) some C A (1)	**all C A (16)**
NVC (1); Err Conc (1)	*NVC (2); Err Conc (2)*

Table 9: Valid conclusions in both Tasks for vp1 problems.

All B A **Some C B**	All B A **Some B C**
BAC (5) **CBA (13)** ACB (1) BCA (1)	BAC (9) **BCA (8)** ABC (3)
NVC (2); Err Ind (0)	*NVC (1); Err Ind (1)*
some A C (3) **some C A (17)**	some A C (15) **some C A (5)**
NVC (1); Err Conc (0)	*NVC (4); Err Conc (0)*
Some A B All B C	**Some B A** All B C
ABC (17) BCA (4)	**BAC (9)** BCA (10) CAB (1) CBA (1)
NVC (1); Err Ind (0)	*NVC (1); Err Ind (0)*
some A C (18) some C A (2)	**some A C (6)** some C A (16)
NVC (1); Err Conc (0)	*NVC (0); Err Conc (1)*

Table 10: Valid conclusions in both Tasks for vp3 problems.

All A B	
None C B	
AB-C (12)	-CBA (1)
	C-B-A (4)
	C-A-B (2)
NVC (2); Err Ind (1)	
none A C (5)	none C A (10)
NVC (9); Err Conc (0)	

All A B	
None B C	
AB-C (19)	
BA-C (1)	B-CA (1)
NVC (0); Err Ind (0)	
none A C (16)	none C A (4)
NVC (3); Err Conc (0)	

None A B	
All C B	
A-B-C (3)	
-ABC (1)	**CB-A (12)**
B-AC (1)	BC-A (2)
NVC (2); Err Ind (0)	
none A C (15)	none C A (2)
NVC (7); Err Conc (0)	

None B A	
All C B	
-BA-C (1)	
	CB-A (17)
	BC-A (1)
NVC (0); Err Ind (3)	
none A C (2)	none C A (19)
NVC (2); Err Conc (0)	

Table 11: Valid conclusions in both Tasks for vn1 problems. Each premiss is source for a different individual.

None A B	
Some B C	
-ABC (3)	**BC-A (12)**
	CB-A (2)
NVC (2); Err Ind (3)	
	somenot C A (10)
NVC (9); Err Conc (3)	

Some A B	
None B C	
AB-C (17)	
	AB-C (1)
NVC (8); Err Ind (1)	
somenot A C (1)	
NVC (19); Err Conc (4)	

None A B	
Some C B	
-ABC (1)	**CB-A (14)**
NVC (3); Err Ind (2)	
	somenot C A (3)
NVC (14); Err Conc (7)	

Some A B*	
None C B	
AB-C (7)	
NVC (4); Err Ind (0)	
somenot A C (2)	
NVC (17); Err Conc (5)	

None B A	
Some B C	
B-AC (2)	**BC-A (13)**
NVC (2); Err Ind (4)	
	somenot C A (4)
NVC (13); Err Conc (6)	

Some B A	
None B C	
BA-C (13)	B-CA (3)
AB-C (3)	
NVC (3); Err Ind (0)	
somenot A C (2)	
NVC (16); Err Conc (3)	

Table 12: Valid conclusions in both Tasks for vn3 problems. 11 subjects did not receive the problem marked * owing to an error in the materials set.

All B A **Somenot B C**	
BA-C (14)	**B-CA (2)**
AB-C (2)	
NVC (3); Err Ind (1)	
somenot A C (16)	
NVC (6); Err Conc (1)	

Somenot B A All B C	
B-AC (3)	BC-A (15)
	CB-A (1)
NVC (2); Err Ind (1)	
	somenot C A (16)
NVC (6); Err Conc (1)	

Table 13: Valid conclusions in both Tasks for vn4 problems.

All A B **Somenot C B**	
-A-BC (1)	**C-B-A (9)**
	C-A-B (1)
NVC (8); Err Ind (2)	
	somenot C A (7)
NVC (12); Err Conc (5)	

Somenot A B All C B	
A-B-C (5)	-C-BA (2)
NVC (7); Err Ind (7)	
somenot C A (4)	
NVC (13); Err Conc (5)	

Table 14: Valid conclusions in both Tasks for vn5 problems.

None A B **Somenot B C**	
-AB-C (1)	**B-C-A (10)**
-A-CB (1)	
B-A-C (1)	
NVC (7); Err Ind (1)	
NVC (23); Err Conc (0)	

Somenot B A None C B	
B-A-C (11)	B-C-A (2)
NVC (8); Err Ind (0)	
NVC (24); Err Conc (0)	

None B A **Somenot B C**	
B-A-C (5)	**B-C-A (9)**
NVC (4); Err Ind (4)	
NVC (20); Err Conc (3)	

Somenot B A None B C	
B-A-C (10)	B-C-A (5)
	-C-AB (1)
NVC (3); Err Ind (3)	
NVC (22); Err Conc (1)	

Table 15: Valid conclusions in the Individuals Task for vu2 problems. Standard Task has NVC.

B.2 Non-unique source problems

All B A All B C	
BAC (15)	BCA (5)
ACB (1)	
NVC (1); Err Ind (0)	
some A C (4)	
NVC (3); Err Conc (10)	

Table 16: Valid conclusions in both Tasks for vp2 problems.

None A B All B C	
-ABC (2)	BC-A (11)
B-AC (2)	CB-A (1)
NVC (3); Err Ind (2)	
NVC (13); Err Conc (10)	

None B A All B C	
B-AC (1)	BC-A (17)
	CB-A (2)
NVC (1); Err Ind (1)	
	somenot C A (1)
NVC (10); Err Conc (13)	

All B A None C B	
BA-C (12)	-CBA (1)
AB-C (1)	
NVC (5); Err Ind (3)	
NVC (11); Err Conc (11)	

All B A None B C	
BA-C (18)	
AB-C (2)	
NVC (0); Err Ind (2)	
somenot A C (1)	
NVC (6); Err Conc (14)	

Table 17: Valid conclusions in both Tasks for vn2 problems.

None A B None B C	None B A None C B
-AB-C (1) +B-C-A (6) B-A-C (4)	B-A-C (6) B-C-A (1) -CB-A (1)
NVC (6); Err Ind (3)	*NVC (10); Err Ind (2)*
NVC (15); Err Conc (9)	*NVC (18); Err Conc (5)*

None A B None C B	None B A None B C
-AB-C (1) B-C-A (1) B-A-C (8)	-AB-C (1) -A-CB (1) B-A-C (12)
NVC (9); Err Ind (3)	*NVC (4); Err Ind (5)*
NVC (22); Err Conc (2)	*NVC (17); Err Conc (4)*

Table 18: Valid conclusions in the Individuals Task for vu1 problems. Standard Task has NVC.

B.3 Problems without valid conclusions

			Ind. Task		Std. Task	
Diagram	Premiss1	Premiss 2	Err Ind	NVC	Err Conc	NVC
np1	All AB	All CB	11	11	16	7
np2	All AB	Some BC	13	9	20	2
	Some BA	All CB	11	11	18	3
	Some AB	All CB	6	15	16	7
	All AB	Some CB	6	16	13	9
np3	Some AB	Some BC	4	18	6	17
	Some BA	Some CB	4	18	3	20
	Some AB	Some CB	3	18	2	22
	Some BA	Some BC	6	16	3	21
nn1	All AB	Somenot BC	11	11	15	6
	Somenot BA	All CB	12	9	17	4
nn2	Somenot AB	All BC	8	12	10	12
	All BA	Somenot CB	7	15	10	12
nn3	Some AB	Somenot BC	6	16	3	21
	Somenot BA	Some CB	6	16	5	18
	Somenot BA	Some BC	4	18	3	20
	Some BA	Somenot BC	4	18	1	22
nn4	Somenot AB	Some BC	3	18	4	20
	Some BA	Somenot CB	3	19	2	22
	Somenot AB	Some CB	2	19	3	20
	Some AB	Somenot CB	3	19	2	19
nn5	Somenot AB	None BC	6	16	2	22
	None BA	Somenot CB	5	17	2	21
	Somenot AB	None CB	5	16	2	20
	None AB	Somenot CB	3	19	3	21
nn6	Somenot AB	Somenot CB	3	19	1	22
nn7	Somenot AB	Somenot BC	2	20	2	22
	Somenot BA	Somenot CB	3	19	2	21
nn8	Somenot BA	Somenot BC	2	20	1	21

Table 19: Problems without valid conclusions in both Tasks: numbers of invalid conclusions and correct NVC responses.