Image and language in human reasoning: a syllogistic illustration

Keith Stenning & Peter Yule

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

Existing accounts of syllogistic reasoning oppose rule-based and model-based methods. Stenning & Oberlander (1995) show that the latter are isomorphic to well-known graphical methods, when these are correctly interpreted. We here extend these results by showing that equivalent sentential implementations exist, thus revealing that all these theories are members of a family of abstract *individual identification algorithms* variously implemented in diagrams or sentences. This abstract logical analysis suggests a novel *individual identification task* for observing syllogistic reasoning processes. Comparison of the results of this task with the Standard Task confirms that the tasks are psychologically closely related, throwing light on sources of error, on subjects' sensitivity to metalogical properties, and on term ordering phenomena. Since it avoids posing the sub-task of formulating a quantified conclusion, the new task allows comparison of explanations of problem difficulty in terms of the number of models (e.g. Johnson-Laird & Bara 1984) with alternatives in terms of the difficulty of choosing a quantifier for the conclusion. Logical concepts of *source* and *conditional* premisses provide a comprehensive account of term order data, including figural effects, at a level abstract with regard to imagistic or sentential representations. These results argue that much richer empirical evidence will be required to discriminate phenomenologically distinct reasoning processes than has hitherto been supposed.

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: logics are abstract consequence relations;[1] only implementations, of logics or of fragments of logics, are *mechanisms*. The relation between logics and implementations is many-levelled and indirect. If this is so for machines, we should expect it to be more so for human beings. To understand the mind's computations we should seek to describe generalisations at the most abstract level at which they hold, but that level will often be far above the mechanisms involved. 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.

This paper applies semantic and computational analysis to the apparently contrasting *external* representations which figure in psychological theories about the *internal* representations engaged in syllogistic reasoning. Its purpose is to show that the important behavioural phenomena observed when people solve syllogisms 'in their head' can be described at a level of abstraction above that which distinguishes the theories that have been proposed. In other words, the computational commonalities of these theories are obscured by their contrasting representations but can be unearthed by logical analysis.

Cognitive theories of mental processes 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).

The imagery debate debates the nature *internal* representations. 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 individual, *token* representations and *systems* of representation by applying the concept of *expressiveness.* They propose that when graphics is 'good' it is because graphical *systems* are inexpressive relative to languages, that is to say, they impose more constraints on the types of information which they can represent. Although *token* linguistic and graphic representations (particular paragraphs and diagrams) may be informationally identical, the linguistic ones are generally drawn from *systems* which are much more expressive (i.e. have tokens for which there are no graphical equivalents) and this is the reason why a general sentential processor cannot be so efficient.

Readers unfamiliar with the computational theory of expressiveness might read Levesque 1988 but perhaps a simple example will suffice to illustrate. Most diagrammatic systems of

[1] Table 15 contains definitions of the logical terms used.

representation force the determination of all identity relations between represented objects. It is hard to draw a picture of 'the morning star and the evening star' without depicting them as one star or alternatively as two. It is all too easy to avoid this determination in describing them in expressive languages such as English. English is more expressive than such diagrammatic systems because it can abstract over both kinds of situation. This expressive power has both advantages and disadvantages. One of the disadvantages of expressiveness is that it generally leads to inferential complexity. Not all languages are expressive. Inexpressive languages (e.g. ones with unique names and no quantifiers) may enforce the determination of all identity relations just as diagrams do. *Specialised* sentential processors for such languages can be just as efficient as graphical ones but general processors will not be.

The major distinction between psychological models of mental representation and process in this field is between sentential ("rule-based") and model-based theories. Stenning & Oberlander (1995) have clarified the 'graphical' nature of the model-based theories, by showing that the Mental Models method (Johnson-Laird 1983) is strongly computationally equivalent to a rational reconstruction of Euler's (1772) (and Leibniz' 1666/1966) method using graphical representations composed of circles.

This does not mean that the two theories are psychologically identical. Reasoners' access to the metalogical properties of the syllogism may be quite different for the two representational systems. The graphical method renders the logical constraints of the domain explicit in the formalism. For example, the graphical relation of inclusion is transitive like the logical relation it represents. But with the semantic net formalism of Mental Models theory, the graphical relations (labelled nodes joined by links) do not have the logical properties of the relations they depict. Such semantic network formalisms are known to be, like English, indefinitely expressive (see Schubert 1976). An immediate example consequence is that while Euler's sytem is *self-consistent*, the mental models notation can perfectly well represent inconsistent sets of propositions. This may make every difference to a learner seeking an efficient method of reasoning in this domain. But it remains the case that as far as reasoning *within* established systems goes, the two representations paired with their respective operations are computationally indistinguishable. The experimental evidence researchers have sought has been about reasoning within systems, and has mostly been simply about input/output patterns of error, conclusion forms, and timings. Researchers have not sought the kind of evidence (for example about learning) that might distinguish these theories.

We now extend this semantic analysis of the syllogism to equivalences between the graphical Euler system and a sentential natural deduction system in order to explore the consequences for explanations of some of the main behavioural phenomena of syllogistic reasoning. 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 the sentential and imagistic representations; suggests novel ways of gaining behavioural evidence about the mental process of syllogistic reasoning; proposes alternative explanations of difficulty and error; and reveals an abstract level at which all the important behavioural generalisations, currently posed in different models, are unified. These demonstrations do not deny that there may be computational differences between reasoning *about* imagistic and *about* sentential implementations. On the contrary, we strongly believe that important individual differences in mental processes hinge on differences in implementation. But these demonstrations do show that much current evidence of reasoning *within* these systems is in principle inadequate to differentiate these models, and raise 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*—only those syllogisms which entail the existence of a fully specified *type* of individual have a valid conclusion. We adopt the term *case identifiability* to suggest this metalogical property, that all conclusions are valid in virtue of an identifiable 'case'. For example, if a, b and c are the terms of a syllogism, then there are eight types of individual defined by their having or not having the three properties denoted by the terms. The premisses of the syllogism *All b are a. Some b are not c. Therefore, some a are not c* entail that there are individuals that are a and b but not c (a maximally specified type), whereas the premisses *All a are b Some b are c* entail no type specified with respect to all three properties and so entail no valid conclusion.

Examination of the converse of this generalisation reveals that there are valid conclusions which follow from syllogisms that do entail fully specified types of individual, but which cannot be expressed with the conventionally allowed quantifiers. We call these *U-conclusions*, and if we extend the syntax of the syllogism to accommodate them, the converse generalisation, *that every syllogism which entails a fully specified type of individual has a valid conclusion*, then holds (see problems type vu1 and vu2 in Figure 9).

Because the syllogism is case-identifiable, there is one abstract family of algorithms (which we call *individual identification algorithms*) which characterise a fully specified type of individual, and then draw quantified conclusions by abstracting away from this characterisation. Whilst there may be other algorithms for syllogism solution, we claim that all extant psychological theories of how people solve syllogisms fall into this abstract class.

This demonstration affords unique opportunities to recast models of syllogistic reasoning at a more abstract level and to collect novel data about the mental processes concerned. A novel abstract description reveals new behavioural generalisations. We raise the question whether explanations of these generalisations need to appeal to differences between imagistic or sentential representations, or whether the key behavioural observations can better be accounted for at a level of abstraction above these implementational issues. We present analyses which argue for the latter answer.

There are two chief kinds of data which previous theories have modelled—errors in reasoning, and *figural effects* on 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 subject/predicate structure of conclusions, making it possible to identify figural effects [2]. 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), Johnson-Laird & Byrne (1991)).

With regard to reasoning error patterns, Mental Models theory's predictions derive from the number of models the theory requires a reasoner to construct for a problem. The more models the more likely a reasoner will fail to search or find all the cases. Johnson-Laird & Bara (1984) showed that their distinction between three-model and two-model problems was implementation-dependent, and in more recent work (e.g. Johnson-Laird & Byrne 1991)

[2] The reader should beware that some authors use the term "figural effect" to refer to the effects of figure on reasoning accuracy, whereas we use it here to refer to the effects of figure on term order in conclusions.

Mental Models theorists have taken the main distinction to be that between single-model problems and multiple-model problems. Ford (1994) has argued that differences in difficulty between syllogisms owe more to the difficulty of choosing the quantifier of a conclusion than to the number of models: problems in which the quantifier of the valid conclusion is not the same as the quantifier of either premiss are especially difficult, whether single-model or multiple-model, but since all but one of them are multiple-model problems, it is hard to unconfound the two variables. Our novel framework generates new data which bears on this issue.

With regard to figural effects, the important behavioural phenomenon 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 validly 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 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, or to a population of subjects not given to engaging the logical task.

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. All theories, whether based on graphical or sentential representations, require representation of ordered term structures. Such structures can readily be represented as verbal lists or spatial arrays. Why jump to 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. Details of memory implementation will not explain these interactions. What is needed is an explanation of how the grammatical structuring results 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.

The plan of the paper is as follows. 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. 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 sentential 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 this analysis, we present an experiment which uses a novel *Individual Identification Task* in order to: seek a characterisation of the main reasoning phenomena at a level of abstraction above that which distinguishes current models; seek data about subjects' ability to draw U-conclusions (the newly identified conclusions inaccessible in the conventional task); and to explore the sequencing of properties in subjects'

	diagonal		symmetric	
Scholastic no.	4	1	2	3
Johnson-Laird no.	1	2	3	4
Term	a-b	b-a	a-b	b-a
Arrangement	b-c	c-b	c-b	b-c

Table 1: The four figures of the syllogism, with two conventional numbering schemes.

mental processes. The Standard Task condition of the experiment is designed to show that the new Individuals Task is a psychologically as well as logically close relative of the Standard Task, and that the new results are therefore not artefacts of a novel experimental procedure.

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 computationally does set imagistic and linguistic methods apart, and what sorts of psychological evidence might distinguish them.

2 The logical structure of the syllogism

In this section we first describe the structure of the syllogism, and then its semantics so that we can explain case-identifiability. We then show how various algorithms are only applicable because of this metalogical property.

2.1 Model Theory

Syntax There are 4 *moods* or premiss types, distinguished by the quantifiers (and conventional mnemonics) “all” (abbreviated A), “some” (I), “none” (E) and “some…not” (O). The quantifiers *all* and *none* are *universal*. The quantifiers *some* and *some…not* are *existential* or *particular*.

A syllogism consists of two premisses which relate three terms (a, b and c), one of which (the *middle* term, b) occurs in both premisses, while the other two (the *end* terms, a and c) each occur in only one premiss—a is the end term in the first premiss, and c is the end term in the second premiss. Thus, for example, we can notate the premiss *Some a are b* as Iab.

There are four possible arrangements of terms in the two premisses, known as *figures*, as shown in Table 1. There are two commonly-employed numbering schemes for figures in the syllogistic reasoning literature: the Scholastic scheme, deriving ultimately from Aristotle, and Johnson-Laird's (e.g. 1983) scheme. This can lead to confusion, and in this paper we attempt where possible to avoid the use of numbers to denote figures; where this is impossible, we follow Johnson-Laird. We also make use of the term *diagonal figures* to refer to the first pair of figures (ab/bc and ba/cb) and *symmetric figures* to refer to the second pair (ab/cb and ba/bc).

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. The abbreviation AO4 stands for the syllogism *All b are a, Some b are not c* (AbaObc), which has the valid conclusion *Some a are not c* (Oac).

Semantics The semantics of the syllogism concerns *types* of individuals defined by combinations of properties. Because there is no identity relation, it is never of logical significance how many of a type of individual exist—hence the talk of *types*. Properties are based on the terms of the syllogism—given the term a, any individual is either an a or not an a (¬a). Fully specified types are specified with respect to all three properties, and there are eight such types: abc, ab¬c, a¬bc, a¬b¬c, ¬abc, ¬ab¬c, ¬a¬bc, and ¬a¬b¬c.

We can express the semantics of syllogisms in terms of sets of types. *Models* of syllogisms are sets of fully specified types whose existence is consistent with the truth of both premisses. Each set can be thought of as a possible syllogistic world described at the level of abstraction with which the syllogism deals. If a type of individual occurs in all possible models of a syllogism, then the syllogism entails the existence of individuals of that type.

Conventionally, the syllogism is interpreted under the assumption that none of the three sets of things with the properties a, b and c are empty. Subjects naturally adopt this assumption under the instructions used in the experiments in the literature (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. A modern logical treatment would deal with these existential assumptions separately from its definition of the quantifiers.

Inferential Structure Of the 64 syllogisms 27 have 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 fully specified type of individual.*[3] So, for example, *Some b are not a. All b are c.* entails that there are individuals of the fully specified type ¬abc (and therefore that *Some c are not a*). There is a direct relation between the entailment of the existence of such types and the entailment of existential conclusions—merely removing the middle term from the specification yields the existential conclusion. What will concern us later is that the existence of the maximal individual is semantically richer than the conventional existential conclusion. It is this richness which is exploited in the experiment below. Table 13 in Appendix A.1 shows the necessary individuals for each of the 64 premiss pairs; some readers may prefer Figures 9 and 10 in Appendix A.2 for a more perspicuous graphical presentation of this information.

Considering the converse of the above generalisation is revealing. It turns out that for all but one small group of syllogisms, *if they entail the existence of a fully specified type, they also have a valid conventional conclusion.* The exceptions are a group of syllogisms with two negative premisses. For example, *No a are b. No b are c* establishes that there are ¬ab¬c individuals. The quantificational apparatus of the conventional syllogism will not allow expression of the conclusion *Some ¬a are ¬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 mentioned above). One of the two main generalisations of Aristotle's

[3]This generalisation can be confirmed in Figures 9 and 10 in Appendix A.2 by observing that all registration diagrams of syllogisms with valid conclusions contain a × (see the presentation of the graphical method in Subsection 2.2.1).

meta-theory of the syllogism was that two negative premisses had no valid conclusion. 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. The importance of this distinction was not fully understood until the logical work of Russell in the early 20th century.

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 *Something is a or b. Something is a or c. Nothing is a. Therefore something is b or c.* are not case identifiable. These premisses are satisfied by the model consisting of the following set of types: {¬abc, ¬ab¬c, ¬a¬bc}, and several of its subsets(i.e. {¬abc}, {¬ab¬c, ¬a¬bc}, {¬abc, ¬ab¬c}, and {¬abc, ¬a¬bc}) There is no type which belongs to all of these subsets. So none of the particular types are entailed to exist by the premisses. Therefore 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. The graphical methods described below cannot represent partially specified types and so rely on the fact there is no need to do so for syllogisms because they are case-identifiable. The formulation of the sentential algorithm given below is also dependent on the case-identifiability property.

2.2 Solving syllogisms by identifying critical individuals

Our purpose in this section is to describe an abstract algorithm for syllogistic reasoning, the individual identification algorithm. We 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.

2.2.1 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 redescribe the algorithm here for the reader's convenience.

Premiss representation In the graphical implementation of the method, individual types are represented as regions in a diagram defined by circles (see Figure 1). The *maximal* model of a premiss (the largest possible model of types consistent with the truth of the 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 minimal model (the smallest possible model of types consistent with the truth of the premiss, i.e. the set of types whose existence is entailed by the premiss) is represented by marking with an "×" the regions of this diagram which correspond to types of individuals which *must* exist if the premiss is true. We call these marked regions *critical regions*, and they represent *critical individuals*. Figure 1 shows the 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—so they can be resized and moved at will, provided no regions are created or abolished. This premiss representation is a direct translation of the set-theoretic relations into graphical terms.

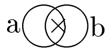
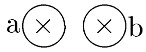
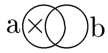
All a are b		A
Some a are b		I
No a are b		E
Some a are not b		O

Figure 1: Euler diagrams for each premiss type. Regions representing necessary individuals are marked with a ×.

Case Identification A new diagram is constructed by 'registering' two premiss diagrams by superimposing the middle term circles. There may be more than one way of doing this consistent with the constraints imposed by the premiss diagrams, 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 is possible without violating the topological constraints imposed by the premiss diagrams. The result is the *registration diagram*, which represents the maximal model of the premiss pair. Figure 2 shows an example; in this case the centre registration diagram is the maximal one. Now each critical region deriving from the original premiss diagrams will either have been bisected by the third circle or not; if the critical region is bisected, then the mark is removed.

The attempt to overlap the end-term circles ensures that all possible fully specified types are represented by the resulting diagram. So if a critical region is not bisected, and consequently remains marked in the registration diagram, the corresponding fully specified type must exist. In the example it is the type ab¬c.

These diagrams allow the concept of *source* premiss to be defined. Any × mark persisting in a completed registration diagram can be traced to one or more of the premisses as its source. In the case in Figure 2, the mark comes from both premisses, and so both are possible source premisses. In most syllogisms, only one premiss can be the source. Figure 4 provides an example.

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 a are b* ↔ *No b are a*) means that the classification abstracts over these differences too. Figures 9 and 10 in Appendix A.2 show the resulting set of 21 distinct registration diagrams, for problems with, and problems without, valid conclusions respectively.

Existential conclusions can be drawn from the descriptions of critical individuals (which are essentially existentially quantified conjunctions of properties) by merely removing the middle term. Universal conclusions are only warranted if the critical region is circular. If so, its label becomes the subject term of a universal quantification.

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 exhibited by some methods which are ostensibly based on Euler Circles (e.g. Erickson 1974, Guyote & Sternberg 1981). Note also that these properties are only possible because the syllogism is case identifiable (we return to this point below).

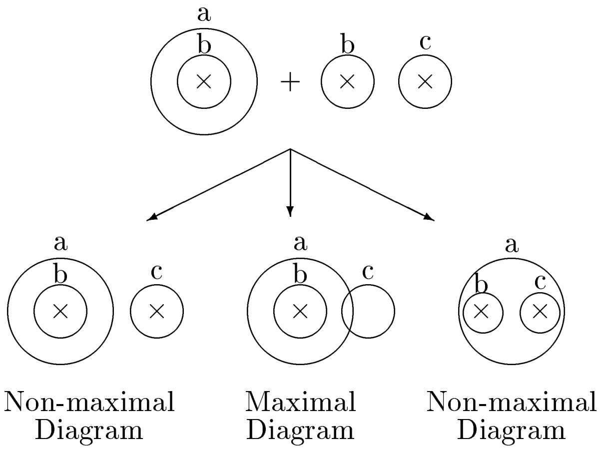


Figure 2: Constructing a registration diagram for the example *All b are a, No b are c.* While any of the three possible registration diagrams is consistent with the premisses, the central one should be chosen since it has the largest number of regions.

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	$A \& \neg B$

Table 2: Propositional representations for each premiss type.

2.2.2 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 underlying propositional operations. Under this interpretation, syllogistic premisses are encoded into propositional calculus according to the scheme in Table 2[4].

The rules of inference &-elimination, &-introduction, $\rightarrow$ elimination, (otherwise known as *modus ponens*), double negation, and *modus tollens* are sufficient to draw all valid syllogistic conclusions (most introductory logic books give an account of rule-based natural deduction systems, though the reader will not require any further knowledge to understand the small fragment used here). Within this propositional system, we can conceive of the business of drawing syllogistic conclusions as the process of building up descriptions of single critical individuals. 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 of this individual [5].

[4]Capital letters are used in conformance with the usual propositional calculus conventions. See e.g. Lemmon (1965) for an introduction to propositional calculus.

[5]It is this focus on a single individual, licensed by case-identifiability, which permits this propositional formulation.

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 modus ponens. From B together with the second premiss, we derive $\neg C$ by modus tollens. We can now construct the specification of a critical individual: $AB\neg C$ (by conjoining the terms with &-introduction). Having characterised a critical individual, conventional existential syllogistic conclusions can then be drawn by eliminating the middle term (by &-elimination). Table 14 in the Appendix A.3 presents the propositional representation, the initial assumption, and the rule used for each valid syllogism.

Table 3 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 individual description; completing the individual description with information from the conditional premiss by applying modus ponens or modus tollens; and drawing abstract conclusions from the completed individual description.

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 modus ponens or modus tollens, to conclude the third conjunct of the individual description. Finally the third part abstracts conclusions from the individual description by eliminating the B conjunct and introducing a quantifier. This abstract description makes it clear that this algorithm is a member of the class of individual identification algorithms. Case identifiability is reflected in the construction of fully specified descriptions of individuals.

2.2.3 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 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 all circles are positive. Premisses with any $\times$s outside any circles are negative. The subject term of the sentential form always corresponds to an $\times$-marked region.

Figure 3 provides a graphical expression of the sentential rules *modus ponens* and *modus tollens*. The central column contains representations of the conditional premiss. The two premiss diagrams with the b circle appearing as a critical region represent the only two premisses in which b is an antecedent of a conditional, and therefore the only two premisses which function as conditionals in modus ponens. In the left column are the five diagrams in which the b circle contains an $\times$, corresponding to the five premisses which entail b, the other premiss required to match the antecedent of the conditional 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. The graphical operation of registering the b circles therefore corresponds to sententially concluding either the consequent or its negation (as the

1. Seek a unique existential premiss:
    (a) If there are two, then respond NVC and quit.
    (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 individual description. If the source premiss is universal, assume its antecedent. Apply modus ponens and conjoin the consequent to the antecedent to make the first two clauses of the individual description.
4. Compare middle terms:
    (a) If a source middle term matches (with regard to negation) the antecedent middle term of the conditional premiss, apply modus ponens, and conjoin consequent term to individual description. Go to 5.
    (b) If the source middle term mismatches (with regard to negation) with the conditional consequent middle term, apply modus tollens to the conditional premiss, and conjoin consequent term to individual description. Go to 5.
    (c) ELSE conclude NVC and quit.
5. individual description is now complete.
6. Draw Abstract conclusion from individual description:
    (a) Delete B conjunct from individual description. 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.

Table 3: The 'sentential' algorithm based on the process of constructing an *individual description* by conjoining terms for each predicate or its negation ('no valid conclusion' is abbreviated NVC)

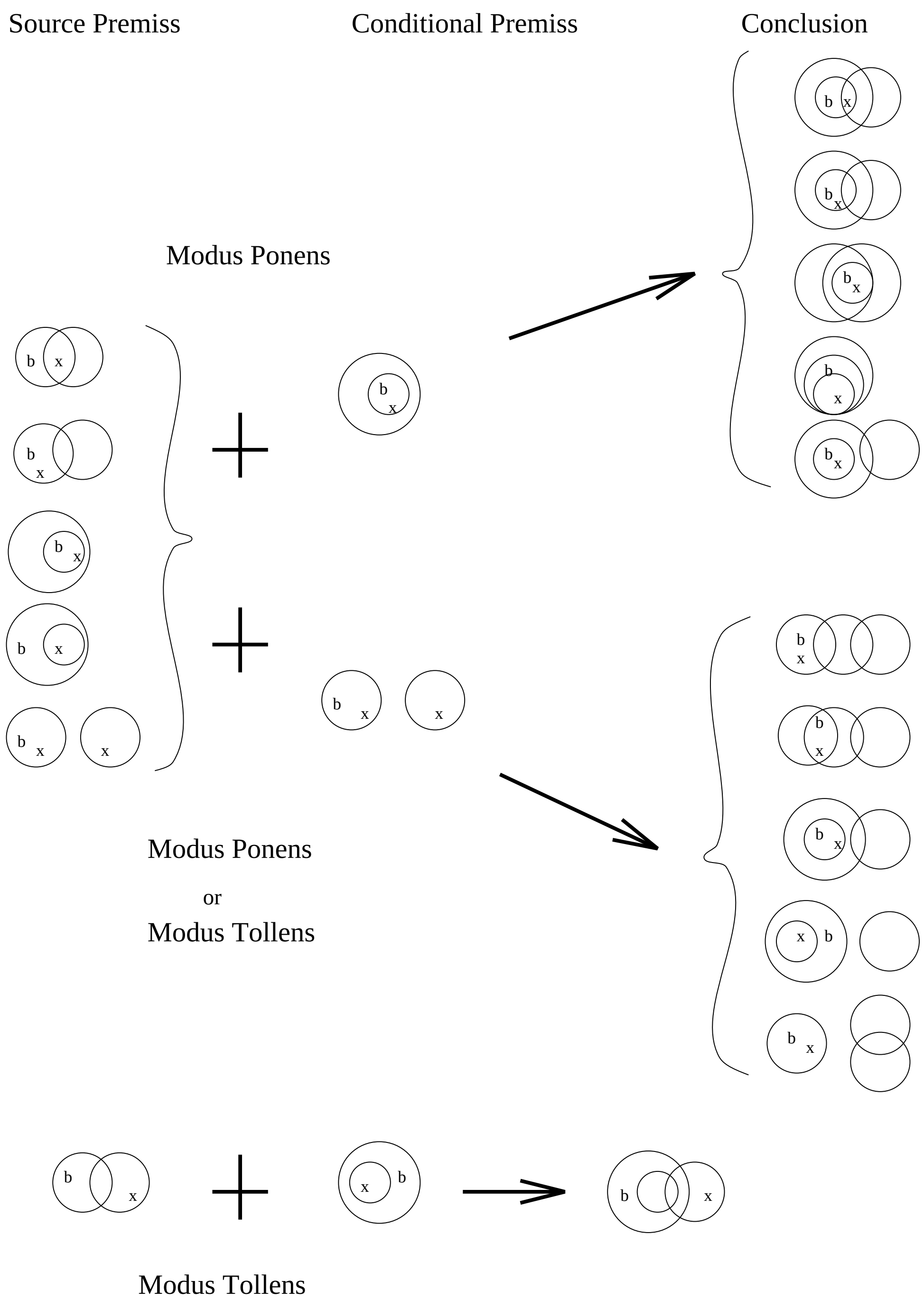


Figure 3: Graphical correspondences to sentential rules: the 5 diagrams in which b contains an $\times$ are registered with either of the two universals with b as a critical region in the analogue of *modus ponens*. The one diagram in which b contains no $\times$ is registered with the one universal with b as a non-critical region in the analogue of *modus tollens*. The two rules yield the 10 diagram-types with valid conclusions in Figure 9

non-source circle includes or excludes the critical region). Its inclusion in, or exclusion from, the end-term circles is always continued in the final diagram.

Although universal negatives can be represented as $A \to \neg B$ they can also be represented as $B \to \neg A$. So for each application of modus ponens to a universal negative conditional, there is a corresponding application of modus tollens which has the same result. Hence, in the middle row of the table, the label *modus ponens or modus tollens*. Without more information about the particular representation that is employed we cannot tell which rule is used.

In the bottom row, 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. 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, or exclusion from, the end-term circles is continued in the final diagram. This graphical operation of registering the b circles therefore corresponds to the sentential operation of concluding the negation of the antecedent and conjoining it to the rest of the description from the source premiss. The resulting 10 types of registration diagram in the right column (the fifth and eighth from the top are of the same type) correspond to all the registration diagrams licencing valid conclusions.

But there is a general difference between the graphical and sentential algorithms as they have been stated thus far. The graphical algorithm completes the registration procedure fully determining which sub-regions appear in the final diagram before reading off any conclusion. The sentential algorithm is focussed on the construction of the individual description 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 proves or disproves the individual conclusion to which it corresponds before deciding any issues related to other regions in the final diagram. Put another way, the graphical algorithm tests all critical individuals in parallel; 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—it is trivial to modify the graphical algorithm to make it focussed, by constructing separate registration diagrams for each critical region, to test each critical region separately. When Euler's method is conceived this way, the graphical and sentential algorithms coincide completely.[6] Even when our phenomenological experience may be of manipulating mental circles, it is a moot point whether all the subregions of the more complex diagrams are fully differentiated. The mental process may be computationally more like the serial process of the sentential method than the parallel process of the physical graphical method.

This presentation emphasises several generalisations about syllogistic reasoning. At their core, syllogisms are conditional inferences. Both methods embody the same functional distinction between the two premiss representations: for any individual conclusion, one premiss supplies the conditional and the other the source of the atomic input to the conditional rule. We return to this distinction 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 graphical and sentential implementations of this algorithm. Case identifiability plays a central role in determining the applicability of

[6]A `prolog` program closely related to the focussed graphical method is described by Yule (1996b).

the 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 the graphical algorithm can be recast in this more focussed way.

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. We have seen there are two main strands of evidence: error patterns and figural effects. Our identification of the individual identification algorithm suggests a way of generating richer empirical data bearing on these issues.

3.1 The Individuals Task

The individual identification algorithm is a method for deciding whether there is any fully specified individual which must exist in any model of the premisses, and of constructing a specification if any does exist. The task which most directly operationalises this process is the task of giving descriptions of fully specified individuals whose existence is entailed by the premisses. Rather than ask subjects to produce conclusions linking the end terms, in one of the four moods A, I, E and O (which we call the *Standard Task*), instead they are asked to decide whether there is any type of individual which must exist given the truth of the premisses, and if so to describe it. We call this the *Individuals Task*. Individual descriptions must make use of all three terms in the syllogism, and each term may optionally be negated (e.g. "an artist who is a beekeeper but not a chemist" or "artist not chemist beekeeper").

This task has several major differences from the Standard Task—there is no overt quantification of conclusions in the Individuals Task, and there are more syllogisms with valid conclusions, since the U-conclusions are valid. But Individuals Task conclusions also have term-ordering properties which give the Individuals Task several advantages 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—so term order in conclusions is free. By contrast, in the Standard Task not all conclusions are validly convertible (I and E can be converted; A and O cannot), and so the sequence of terms is to some extent confounded with validity of conclusion. This problem is compounded by the fact that in the diagonal figures, valid conclusions also tend to be figural ones, so any perfect performance in the Standard Task would always be expected to show a figural effect, irrespective of the reasoning method used. If we use Standard Task data this problem both restricts the empirical base for measurements of any psychological or implementational figural effect, and makes it difficult to treat the logical and psychological figural effects as aspects of a unitary phenomenon.

Another advantage of the Individuals Task is that it yields data on all three terms of the syllogism, rather than just on end terms. Any description can be formulated in six logically equivalent property orders: bac, abc, acb, bca, cba and cab. While we can still look at the

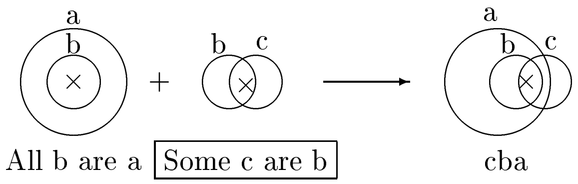


Figure 4: Registration diagram for a problem (type vp3) with source premiss 2.

relative positions of the end terms for comparison with the Standard Task figural effect, we also have data on the position of the middle term. This allows new hypotheses to be tested; in particular Mental Models theory would be expected to make predictions about the position of middle terms in conclusions, since it explains end-term order in conclusions by positions of end term tokens in intermediate representations which also include middle term tokens.

The Individuals Task provides richer data, but our semantic analysis also suggests new and more abstract generalisations about subjects' inferential behaviour. We first elaborate an account of figural effects in Subsection 3.2 in terms of the distinction between source and conditional premisses, and then test predictions in Subsection 3.3 using data from the novel Individuals Task.

3.2 Individual Descriptions and Figural Effects

In Section 2 above we introduced the concepts of source and conditional premisses in the derivation of individual description conclusions. To recap briefly, the source premiss for some valid individual conclusion is any premiss which makes an existential assertion, from which the existence of that individual can be inferred by using the other (*conditional*) premiss as a rule of inference. For each valid individual for each syllogism, that individual may specify a unique source (e.g. see Figure 4), or fail to because more than one derivation of that individual from those premisses is possible (e.g. see Figure 2).

The source/conditional premiss assignment of problems[7] is empirically interesting because it is distributed unevenly among the four figures, in a way that appears to parallel the figural effect. Table 4 shows the numbers of problems with each source/conditional assignment, in each figure, for conventionally valid and U-valid problems separately. For example, in ab/bc, there are 3 conventionally valid problems with source premiss 1, of types[8] vp1, vp3 and vn3, but only one conventionally valid source 2 problem, of type vn3; however there is also a U-valid source 2 problem (vu2), and 3 problems without unique source premisses (vn1[7], vn2 and vu1). In ba/cb this pattern is inverted, since the problems in ba/cb are equivalent to those in ab/bc but for premiss inversion. In the symmetric figures, where premiss inversion does not change figure, each unique-source problem has an inverted counterpart in the same figure, and so no overall imbalance in numbers of problems with each source premiss.

The imbalance in the diagonal figures is related to 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 unconvertible ac valid conclusions

[7]Since most problems entail only one valid individual conclusion, for simplicity of exposition in this section we treat source/conditional assignment as a property of the problem, rather than the conclusion. Consequently vn1 problems, which entail two different individual conclusions, with different sources, are treated as non-unique source problems here.

[8]See Figures 9 and 10 in Appendix A.2 for information on problem types.

	figure			
	ab/bc	ba/cb	ab/cb	ba/bc
Source premiss 1	3	1+U	2	3+U
Source premiss 2	1+U	3	2	3+U
No unique source	2+U	2+U	2+U	3+U
Total	6+2U	6+2U	6+U	9+3U

Table 4: Distribution of source/conditional assignments between figures. Numbers of conventionally valid and U-valid problems are given separately.

than ca ones. The distribution of source parallels this logical figural effect, so that, usually, unconvertible ac conclusions have source premiss 1, and unconvertible 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 4, 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 can examine term ordering effects independently of logical effects. The individual identification algorithm does not itself determine the ordering of terms in descriptions. Indeed, the abstract approach neatly separates the logic from this implementational detail. Different implementations of the algorithm can clearly be devised to yield any combination of the six possible term orders for any description. It is equally clear that these implementations will not be differentiated by whether they employ sentential or imagistic representations. It is merely a matter of how properties are 'read off' whatever representation is used in the course of formulating descriptions.

But the algorithm does suggest a strikingly simple psychological hypothesis about why the Standard Task figural effects observed in the literature might result. We hypothesise that in the Individuals Task, *the terms from the source premiss will tend to precede the other end term in conclusions, and will tend to retain the order in which they appear in the source premiss* (Source-Founding Hypothesis). 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 effect as the consequence of a strategy of beginning the construction of the individual description from the source premiss. If this hypothesis is upheld, it offers the possibility of explaining the complex interactions between figure and quantifiers in the determination of the structure of subjects' conclusions. Moreover, this hypothesis is much stronger than conventional explanations of the figural effect, insofar as it makes specific predictions of end term order on a problem-by-problem basis, for unique source problems, within the symmetric as well as the diagonal figures, rather than merely predicting one or other term order in each diagonal figure, and a random 50-50 split in the symmetric figures.

Our argument is twofold. On the one hand we argue that model-based and sentential theories are not distinguishable on the range of evidence at hand. On the other hand we argue that having an abstract description of syllogistic reasoning in terms of source and conditional premisses allows us to analyse term ordering effects in a way which is quite independent of sentential or graphical representations. Although the Individuals Task yields richer data than the Standard Task, we have shown that is closely *logically* related, and in analysing the experimental data we will show that the two tasks are essentially *psychologically* the same save for the process of abstraction from individual descriptions required by the Standard Task. We are distinctively *not* arguing for a third kind of theory for syllogistic reasoning competing with sentential, graphical, or model-based theories. We criticise these theories not as wrong but as at too low a level of abstraction to be fully informative.

3.3 The experiment

The experiment collected reasoning data on performance in the novel Individuals Task to throw new light on the source of errors and on term ordering phenomena in reasoning; 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 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. We predict, on the basis of our logical analysis, that this should be high. The comparison also specifically provides information about the impact of the conclusion quantifier formulation part of the Standard Task which the Individuals Task does not have.

The Individuals Task also provides evidence about subjects' grasp of the U-conclusions. If subjects correctly draw U-conclusions in the Individuals Task, and correctly observe that there is no valid conclusion in the Standard Task, they thereby exhibit a considerable abstract grasp of the logic of the syllogism. This result would be hard to model in terms of specific rules for inference based on specific representations. The task variable would be shown to control the subjects' judgement of validity.

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 the Source-Founding Hypothesis introduced above, which predicts 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 *the end term of the source premiss occurs before the end term of the conditional premiss.* 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 (type vn1). These have an A premiss whose subject is an end term, and an E premiss (e.g. AabEbc - Figure 5 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. ab¬c supports Eac (or Oac) and c¬b¬a supports Eca (or Oca)). Since

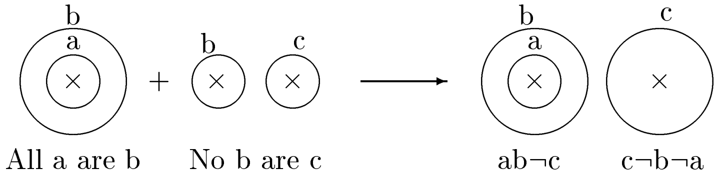


Figure 5: Registration diagram for a problem (type vn1) with two valid individual conclusions, each with a different source.

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.[9]

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 overall 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 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 other 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.

3.3.1 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 took part in the Individuals Task condition, and 25 Edinburgh University students took part in the Standard Task condition. None had any prior training in syllogistic logic.

Materials Each subject received a booklet containing each of the 64 distinct problem types, each printed on a separate slip of paper, in random order. The vocabulary items were selected from sets of nouns denoting nationalities, professions and interests, such that for each problem, each of the three terms was selected from a different set, for example

[9]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.

```
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. The full set of vocabulary items is listed in Table 17 in Appendix B.1.

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, in one of the four syllogistic moods, and 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.

3.3.2 Results

Reasoning accuracy: comparison between tasks In order to investigate the relation in performance between the Individuals Task and performance on the Standard Task, two scattergrams were constructed, showing data from problems without valid conclusions, shown in Figure 6 (see Tables 19 and 21 in Appendix B.2 for data), and with valid conclusions, shown in Figure 7 (see Tables 18 and 20 in Appendix B.2 for data). Each point in each scattergram represents a single problem, and its values on the x and y axes respectively represent the percentages of subjects who responded correctly to it on the Individuals and Standard Tasks. Each problem is labelled with the name of its problem type as explained above (see Figures 9 and 10 in Appendix A.2 for full listings of problem types with and without valid conclusions respectively).

Figure 6 shows the data for problems without valid conclusions (NVC problems). Although scores on the Individuals Task are higher than those on the Standard Task, problem difficulty in the two tasks is quite highly correlated ($r = 0.63,\ 27\ d.f.,\ p < 0.0005$).

While there is no clear separation of groups of problem types in the graph, there is a group of low-scoring problems at the bottom left, of types np1, nn1 and nn2, along with two of the np2 problems, and a group of high-scoring problems, of types np3, nn3, nn4, nn5, nn6, nn7, nn8, along with the other two np2 problems, at the upper right. Although some np2 problems appear in both groups, they appear only in the lower-scoring region of the second group, so np2 might reasonably be classed with np1, np2 and nn1 as a low-scoring problem type. This division of problem types is interesting because the proposed low-scoring group of problem types comprises all and only those NVC problems with at least one A premiss. On the one hand these problems can be picked out as those susceptible to errors due to 'Illicit Conversion' of A premisses (Chapman & Chapman 1959). On the other hand, they may equivalently be described as those NVC problems which *cannot* be decided by the Aristotelian generalisations (viz. that no pair of particular premisses has a valid conclusion, and that no pair of negative premisses has a valid conclusion). For the high-scoring problems, the Aristotelian generalisations will correctly decide the solutions on both tasks, although

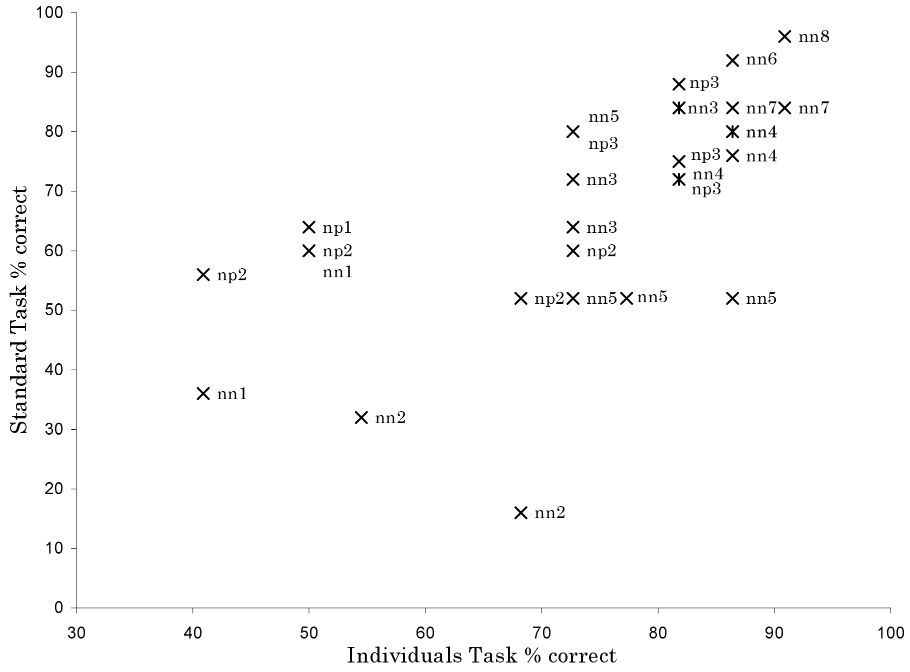


Figure 6: Scatter plot showing scores on NVC problems in both tasks.

there are other problems, the U-conclusion problems, which have valid individual conclusions despite having negative premisses; these are discussed below.

Figure 7 is a plot of performance on both tasks for problems with valid conclusions. Note that the U-conclusion problems do not fall under this definition because they are VC in one task and NVC in the other. But for convenience of comparison, the U-conclusion problems are represented with circles. As it turns out, they appear in the top left region. The valid conclusion results (excluding U-conclusions) are slightly less highly correlated than the NVC results ($r = 0.58,\ 25\ d.f.\ p < 0.005$). As in Figure 6 there is a group of problems with high scores on both tasks: these are instances of diagram types vp1, vp3, and vn1, and as before no other instances of these diagram types appear anywhere else on the chart. The rest of the problems, types vn4, vn3, vn2, vn5 and vp2, are spread out from the middle right to the bottom left.

One might choose to describe this division in terms of number-of-models (as in Mental Models theory), insofar as the first group of problem types (vp1, vp3 and vn1) are single-model problems, whereas with the exception of vp2, the second group of problem types (vn2, vn3, vn4 and vn5) are multiple-model problems.

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-model problems 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

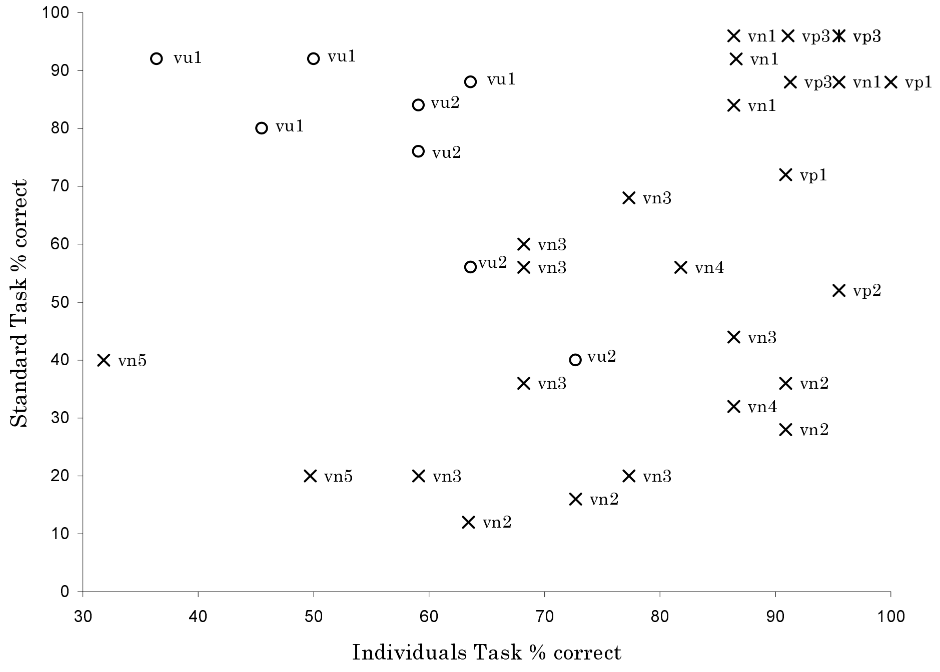


Figure 7: Scatter plot showing scores on VC problems in both tasks.

Individuals Task, which may account for the observed difference in difficulty between tasks for this problem.

In terms of the Individuals Task alone, there is considerable overlap between single-model and multiple-model performance. Problems of type vn4, as well as some vn3 and vn2 problems, show levels of performance on this task similar to that shown by single-model problems. As we have already mentioned, vn2 and vn3 problems have conclusions in a different mood from either of the premisses, so this overlap tends to undermine the explanation in terms of number of Mental Models.

Figure 7 also shows the U-conclusion problems, represented with circles. Note that the Standard Task scores for these problems reflect the percentage of subjects who correctly responded that there was No Valid Conclusion, whereas the Individuals Task scores reflect the number of subjects who correctly drew the valid individual conclusion. These problems' generally high Standard Task scores are comparable with those of the high-scoring group of NVC problems, whereas their relatively low Individuals Task scores are comparable with those of the other low-scoring valid Individuals Task problems. This pattern of results is compatible with the scenario in which a subset of subjects use the Aristotelian generalisations in the Individuals Task (see above).

For the reason noted by Lee and mentioned above, we take number of models as a general measure of complexity of inference and use it to assess similarities between results from the two tasks. We 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 percentages of error conclusion responses and error NVC responses respectively, as the dependent variables. Table 5 shows for both

	Error conclusion		*Error NVC*	
	Inds Task	Std Task	Inds Task	Std Task
single-model	2.07%	8.00%	4.55%	5.09%
multiple-model	9.38%	24.75%	16.19%	35.75%
U-multiple-model	11.98%		29.55%	

Table 5: Mean Percentages of error conclusions and error NVC responses in both the Individuals Task (N=22) and Standard Task (N=40), for single-model, multiple-model and U-multiple-model problems (Individuals Task only).

Tasks, how mean percentages of error conclusions and error NVC responses vary between single-model, multiple-model and U-multiple-model problems.

Percentages of error conclusions are higher in multiple-model problems than in single-model problems, for both Tasks ($Overall: \ F(1,45) = 40.75, \ p < 0.0001$). There is also a main effect of Task, such that there are more error conclusions in the Standard Task than in the Individuals Task ($F(1,45) = 16.11, \ p = 0.0002$), and there is a significant interaction between number of models and Task ($F(1,45) = 6.27, \ p = 0.0159$), such that there is a greater increase in rates of invalid conclusions, from single to multiple model problems, for the Standard Task than for the Individuals Task. However there is no significant difference in percentages of Error Individuals between multiple-model and U-multiple-model problems ($t(21) = 0.73, \ n.s.$).

Similarly for Error NVC responses, there are main effects both of number of models ($F(1,45) = 95.97, \ p < 0.0001$) and of Task ($F(1,45) = 7.51, \ p = 0.0088$) and there is a Task by number of models interaction ($F(1,45) = 19.38, \ p < 0.0001$). As with invalid conclusions, there is a greater increase in the error NVC rate, from single to multiple model problems, for Standard Task than for Individuals Task problems. Moreover, in the Individuals Task there are significantly more error NVC responses to U-multiple-model problems than to conventional multiple-model problems ($t(21) = 2.66, \ p < 0.01$).

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-model problems cause more errors of both types than do single-model problems. The U-multiple-model problems, which only occur in the Individuals Task, differ from conventional multiple-model 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 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.

The Figural Effect: comparison between 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. We predict that Individuals Task conclusions should show a figural effect similar to that found in typical Standard Task datasets. Table 6 shows mean percentages of conclusions in which the a term preceded the c term (ac conclusions), for each of the four Figures, for the Individuals Task and Standard Task (also shown graphically in Figure 8). It is clear that both Tasks show figural

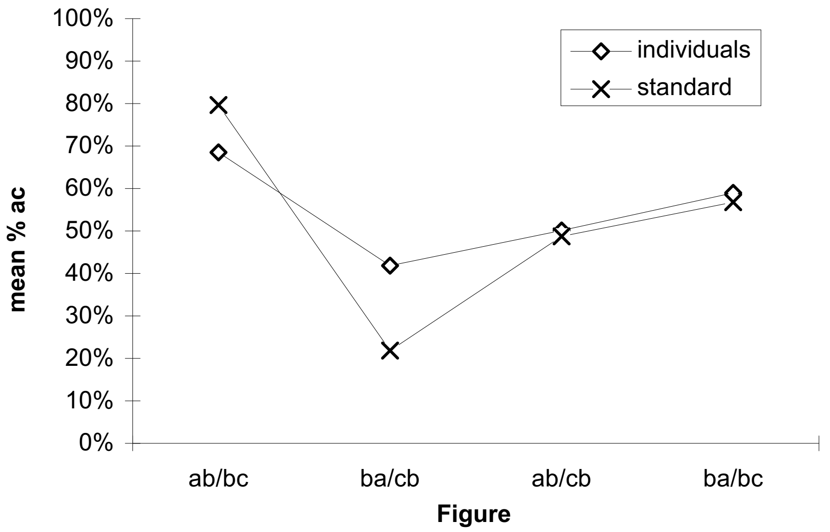


Figure 8: The overall figural effect in the Individuals and Standard Tasks.

		Figure				
Task	N	ab/bc	ba/cb	ab/cb	ba/bc	Overall
Individuals	22	68.53%	41.84%	50.13%	58.96%	54.86%
Standard	25	79.67%	21.85%	48.72%	56.76%	51.75%

Table 6: Mean percentages of ac conclusions in each Figure and Task.

effects, but that of the Individuals Task appears smaller than that of the Standard Task. Also note that in both Tasks, ac conclusions are slightly commoner than ca ones overall.

A mixed-model ANOVA using the independent variables Figure (repeated-measures) and Task (between-groups), with percentage ac conclusions as dependent variable, showed that Figure significantly affected percentage ac conclusions ($F(3, 135) = 44.28,\ p < 0.0001$), but that Task did not ($F(1, 45) = 0.66,\ n.s.$). However there was a significant Task $\times$ Figure interaction ($F(3, 135) = 5.86,\ Greenhouse-Geisser\ adjusted\, p = 0.0015$). Planned comparisons revealed the expected highly significant difference in percentage ac conclusions between the Diagonal Figures (ab/bc and ba/cb) ($F(1, 45) = 104.71,\ p < 0.0001$), which interacted with Task ($F(1, 45) = 14.20,\ p = 0.0005$), and a smaller difference between the Symmetric Figures (ab/cb and ba/bc) ($F(1, 45) = 4.71,\ p = 0.0347$), with no significant interaction between that effect and Task ($F(1, 45) = 0.01,\ n.s.$).

Individuals Task source effects Table 7 shows how end-term order varies with source premiss. The Table shows mean percentage ac order for problems with unique source (Premiss 1 or Premiss 2), and problems in which both premisses are source. Percentage ac conclusions is strongly affected by source premiss ($F(2, 42) = 52.92,\ p < 0.0001$), such that the percentage of ac conclusions is highest when Premiss 1 is the source, lowest when Premiss 2 is the source,

	source			
	Premiss 1	Premiss 2	Both	*Overall*
mean ac % (N=22)	75.95%	30.61%	64.45%	57.00%

Table 7: The Source Effect in the Individuals Task: mean percentages of valid conclusions with ac order in first- and second-premiss unique source problems, and non-unique source problems.

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

Table 8: Frequencies of AE and EA end-term orders for conclusions to two-individual problems (diagram type vn1), labelled A or E depending on the source premiss for the conclusion. End-term order is AE if the end term from the A premiss precedes that from the E premiss, and EA otherwise.

and takes an intermediate value when both premisses could be the source, as predicted by the source-founding hypothesis.

Table 8 shows the relation between end-term order and source premiss for the set of problems which entail two different individual conclusions (diagram type vn1 - see Figure 5). The conclusions can be distinguished as the A-source and E-source conclusions by their semantics alone, and the source-founding hypothesis predicts that in both types of conclusion, the end-term from the source premiss should precede the end-term from the conditional premiss. As the Table shows, 64/68 A-source conclusions have the end term from the A premiss first, and all 10 E-source conclusions have the end term from the E premiss first.

The above analyses cannot distinguish between the effects of source-founding and figure, so it is useful to break down the effects of source founding by figure. Table 9 shows numbers of first-premiss and second-premiss source conclusions in each figure, and percentages of these which have end-term orders as predicted by the source-founding hypothesis. In the diagonal (ab/bc and ba/cb) figures, the source-founding hypothesis predicts end-term order very well, with more than 90% of first-premiss source conclusions having ac end-term order, and second-premiss source predicts ca conclusions in these figures slightly less well, but still much better than chance. Possibly this difference is attributable to the small overall preference for ac conclusions. Note that Mental Models Theory would not predict the Source 2 ca conclusions in ab/bc or the Source 1 ac conclusions in ba/cb for the Individuals Task. Moreover in ab/cb, source-founding predicts end-term order in more than 90% of cases, whereas Mental Models Theory would expect no systematic differences between types of problems in this figure. Finally, however, in ba/bc the source-founding hypothesis fails to predict end-term order at better than the 50% chance level.

The last row of Table 9 shows the apparent figural effect for these data, derived as a function of the frequencies and percentages in the rows above. This illustrates how it is essentially the imbalance in frequencies of conclusions with each possible source premiss,

		figure			
		ab/bc	ba/cb	ab/cb	ba/bc
source 1	ac concs	93.8%	92.6%	90.0%	50.7%
	N concs	81	27	30	75
source 2	ca concs	80.0%	78.1%	91.7%	47.8%
	N concs	30	73	48	67
Overall ac%		73.8%	41.0%	39.7%	51.4%

Table 9: Percentages of conclusions with end-term order as predicted by source founding, in each figure.

		figure			
		ab/bc	ba/cb	ab/cb	ba/bc
source 1	N Subjs	21/22	14/16	14/16	7/22
	binomial p	0.002	0.002	0.002	*n.s.*
source 2	N Subjs	13/18	16/22	16/18	6/20
	binomial p	0.048	0.026	0.001	*n.s.*

Table 10: Numbers of subjects who produced more than 50% predicted conclusions, for each source premiss in each figure, and one-tailed binomial test significance levels.

coupled with the source-founding effect, that gives rise to the end-term order differences between the Diagonal figures.

An alternative view of the prediction of end-term order by source-founding is given by Table 10, which shows numbers of subjects who show the predicted biases for problems in each level of source and figure. Subjects are taken to be biased as predicted if more han 50% of their conclusions in the cell have end-term order as predicted by source-founding. For the first three figures, for both source 1 and source 2 problems, a majority of subjects are biased as predicted, and binomial tests show that significantly more subjects are biased as predicted than would be expected by chance. However, for the fourth figure ba/bc, both binomial tests show no significant difference from chance.

Individuals Task: the middle term We now turn to the predictions concerning the novel data on the position of the middle term (b) in individual conclusions. The hypothesis that individual conclusions are founded on one of the premisses, followed by the end term of the other premiss, predicts two orders in each figure, whereas Mental Models Theory predicts only abc and cba orders. Table 11 shows the percentages of responses with each possible term order in each figure, with the cells corresponding to orders predicted by premiss-founding in **bold**.

It is clear that the orders predicted by premiss-founding are all common, with the predicted cells containing the highest proportions of responses in each figure, by a considerable margin. As expected by both hypotheses, there are few responses with b in final position in any figure, but contrary to the predictions of Mental Models Theory, responses with b in initial position are common—actually slightly more common than those with b in the middle. It is clear that the premiss-founding predictions fit the observed pattern of results far better than do

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

Table 11: Percentages of Individuals Task conclusions with each possible term order in each figure. Orders predicted by premiss-founding are in **bold**, and orders predicted by Mental Models Theory are $\boxed{\text{boxed}}$.

conditional	*founding*			*figure*			*Overall*
quantifier	*premiss*	ab/bc	ba/cb	ab/cb	ba/bc	*Total*	*Total*
E	source	85.3	93.2	88.2	70.3	83.9	71.2
	conditional	7.4	0.0	5.9	23.4	9.5	21.1
	neither	7.4	6.8	5.9	6.3	6.6	7.7
	N concs	68	59	51	64	242	431
A	source	88.4	56.1	77.8	28.2	55.0	
	conditional	9.3	31.7	11.1	61.5	36.0	
	neither	2.3	12.2	11.1	10.3	9.0	
	N concs	43	41	27	78	189	

Table 12: Individuals Task: percentages of source-founded and conditional-founded unique-source individual conclusions in each figure, broken down by the quantifier of the conditional premiss.

the Mental Models Theory predictions.

Individuals Task: conditional-founding We have shown that there is a tendency for the source premiss end term to precede the end term from the conditional premiss in conclusions, and that most conclusions are founded on one or other premiss, in the sense that the terms from one premiss tend to occur in order, followed by the remaining end term. Both results are consistent with the source-founding hypothesis, but we have seen that prediction of end-term order by source premiss breaks down in ba/bc; therefore there must be a substantial number of conclusions which are founded on the conditional premiss in this figure. To investigate this effect, in Table 12 we tabulate source-founded and conditional-founded conclusions in each figure, broken down by the quantifier of the conditional premiss.

It is clear from the table that conditional-founding is commonest, in ba/bc, when the conditional has an A quantifier, and that a smaller, similar effect occurs in ba/cb. Even when the conditional has an E quantifier, however, there are some conditional-founded conclusions in ba/bc. Overall, percentages of conditional-founded conclusions are substantially higher when the conditional quantifier is A than when it is E.

Table 12 also shows overall percentages of unique-source conclusions which are source-founded, conditional-founded or neither; 71.2% are source-founded, and 21.1% are conditional-founded overall.

3.4 Discussion of the experiment

The new Individuals Task is not only logically but also psychologically similar to the Standard Task. The correlations in problem difficulty between tasks support the claim that the Individuals Task recreates the portion of the processes subjects use in the Standard Task, prior to deletion of the middle term and formulation of a quantified statement. While the Individuals Task shows an effect of the number of models, similar to that found in Standard Task datasets, it is significantly less pronounced; this finding can be explained by the fact that the Individuals Task removes one of the causes of difficulty in the Standard Task, namely the problem of choosing a quantifier for the conclusion. Problems whose valid conclusions are in a different mood from either of the premisses are usually also multiple-model problems, so measurements based on Standard Task data tend to exaggerate the difference attributable to the single/multiple model distinction *per se*.

Performance on NVC problems across the two tasks also reveals considerable sensitivity to the logical structure of the syllogism. We have seen that NVC problems can be divided into high-scoring and low-scoring groups, and that this distinction is consistent with at least two equivalent explanations of error in these problems—that the hard NVC problems are those that are susceptible to illicit conversion errors, or that the hard problems are the ones *not* susceptible to solution by Aristotle's rules that no two particular premisses have a valid conclusion, and that no two negative premisses do.

However, many subjects in the Individuals Task condition drew the valid U-conclusions which violate the second of Aristotle's rules. So there is some evidence that many subjects show a sophisticated sensitivity to the different applicability of these rules in the two tasks. If subjects learn which of the Aristotelian rules apply to the task in hand during the course of the experiment, they achieve this learning without feedback.

We have taken the single/multiple models distinction as a general semantic distinction (rather than one tied to Mental Models theory) for the purposes of the analyses reported here[10] Our reconstruction of the Method of Euler Circles recreates the distinction between classes of problems. In graphical terms it turns on the fact that merely superimposing the middle term circles can leave the registration diagram underdetermined. The normative algorithm specifies that the diagram with the greatest number of regions (the maximal diagram) is the one that should be created, but a plausible cause of error here is the creation of a non-maximal diagram. This could only lead to invalid conclusions if some critical region was inadvertently left unbisected, and not all problems with valid conclusions are susceptible to errors of this type—problems of types vp1, vp2, vp3 and vn1 (Johnson-Laird's single-model problems) are not susceptible, but nor are problems of type vn4, which Mental Models theory classifies as multiple-model problems. We note that vn4 problems are as easy as the other single-model problems in the Individuals Task, and are solved relatively easily even in the Standard Task, as this hypothesis would predict. Failure to consider all critical individuals for these 'multiple-model' problems would also tend to cause increased numbers of NVC responses to such problems, again as observed.

[10] One might question whether Mental Models theory in fact makes the distinction in a principled way, as has been done by several authors.

We also note that the lack of further distinction among multiple-model problems leaves Mental Models theory apparently unable to account for differences in difficulty between NVC problems, all of which are necessarily multiple-model problems by that account. It seems likely that subjects use some intuitive grasp of metalogical principles to make NVC judgements. Be this as it may, our current point is that similar processes appear to operate in both standard and individual tasks.

Despite the lack of syntactic constraints on term order in individual task conclusions, we have observed a figural effect similar to, albeit less pronounced than, the figural effect reliably found in Standard Task experiments. This effect appears to be a result of the tendency for the source premiss end term to precede the conditional premiss end term in conclusions, as predicted. The figural imbalance in numbers of conclusions with each possible end term order does appear to be attributable to the figural imbalance in the frequency distribution of problems with different source/conditional assignments. A substantial majority of unique source conclusions are reported in the orders predicted by the more restrictive source-founding hypothesis, with the source premiss terms, in order at the beginning, followed by the remaining end term.

These results are strong evidence for the view that subjects conceive the Individuals Task in a way that is broadly in accordance with the logical framework of the individual identification algorithm. Moreover, the broad similarity between the Individuals Task and Standard Task figural effects suggests that the same conceptual framework underlies Standard task performance, and that the Standard Task figural effect may be explained in much the same way, after taking account of the syntactic constraints on conclusions in that task.

Our source-founding hypothesis explains figural effects in both tasks, but we have seen that the prediction of term order by the source/conditional distinction breaks down in the ba/bc figure, especially when the conditional premiss has an A quantifier. In such cases, conclusions usually have term orders consistent with 'conditional-founding', that is to say that the terms from the conditional premiss occur first in their original premiss order, followed by the remaining (source) end term.

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 hypothesis 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, and similar to what we have observed as the conditional-founding 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 could explain the phenomenon in all the figures is based on substitution of conditional premiss terms for middle terms in source premisses. This would generate 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. But the substitution model 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 source-founding hypothesis alone, the latter are very rare. Substitution as a general explanation has problems.

The alternative explanation is that figural effects are generated by the balance of source- and conditional-founding, and that this balance is affected by the lack of an end-term subject in the ba/bc figure. The balance is different for conditionals with *all* than for those with *no* in all figures. Yule (1996a) discusses conditional-founding at greater length, arguing that it can be explained as the consequence of an unsound solution strategy.

By allowing the observation of the sequencing of all three terms in conclusions, the results from the Individuals Task reveal an extremely simple generalisation of the figural effects of syllogistic reasoning: almost all conclusions are founded on one or other of the premisses, such that the terms from that premiss occur in order before the remaining end term. The very simplicity of this idea perhaps makes it suprising that theories such as Mental Models which actually hypothesise three-term sequences in their underlying representations do not hypothesise the sequences observed here. They diverge because they do not predict that middle-term subjects of founding premisses will typically remain in the initial position of individual descriptions. The underlying reason is that theories such as Mental Models adopt a specific division of processes between representation construction and inference reading off. Mental Models 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 our more abstract characterisation are that it reveals the fact that the inferences can go on in the construction of the individual description, and that the b terms do not have to be 'brought together' to be deleted—conclusions can be read off from anywhere in the individual description. Describing an algorithm at an abstract level reveals what is implementation detail and what is essential to the algorithm.

4 General Discussion

The semantic structure of the syllogism allows processes of inference which simply construct fully specific descriptions of single types of 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, that 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, can be 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.

Establishing these correspondences between algorithms based on different representations enables us to translate insights from one existing theory into another. For example, the mapping from the sentential algorithm through the graphical one onto the mental models one reveals correspondences between Mental Models explanations of error and those of earlier theories such as the Chapmans' (1959) 'illicit 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 Mental Models theory). Lee (1987) demonstrated a close correspondence between the three model problems and the ones where conversion theory predicts errors. The exact details will depend on the details of specific error mechanisms built into either conversion theory or Mental Models theory.

What do these equivalences mean for the psychology of reasoning? They emphatically do *not* mean that every implementation of a syllogistic reasoner using the individuals algorithm with graphical representations is equivalent to every implementation with sentential ones. Rather, given a graphical implementation we can construct a sentential equivalent. The main impact that this result has on psychological modelling is through what it says about the kind of evidence which might distinguish different implementations. So long as the evidence we have is of mappings between premisses presented and conclusions drawn (inputs to outputs) this evidence will not distinguish the kinds of representations involved. Arguments about whether subjects are using some kind of mental models sytem or some kind of sentential systems cannot be based on input–output mappings of premisses to conclusions. If they are, they are vacuous arguments, because each proposal couched in one kind of representation can be mimicked in the other. At this level there literally is no fact of the matter.

There are two ways forward. One is to focus on issues which are independent of representation, as we have done here. We have used the equivalence results to describe data at a more perspicuous abstract level. We have taken this approach by seeking richer empirical evidence of term ordering and describing subjects' grasp of U-conclusions. This approach reveals generalisations which are unavailable when we focus on representational differences. Term ordering processes do not 'get the middle terms into the middle' in order to perform some specific algebraic operation but rather they tend to keep the first two terms in the order they appear in their premiss. With regard to U-conclusion syllogisms, a good proportion of subjects can suitably adapt their grasp of validity as it changes with task. Both these observations are quite independent of the kinds of representation that are being used to compute these processes.

The view of syllogistic reasoning which emerges at this level of abstraction is one in which the two main problems are 1) to decide which if any premiss to treat as source, and which as conditional, and 2) to formulate a quantified conclusion from the critical individual description (in the standard task). Although in general subjects successfully base their conclusions on the source premiss, there is some tendency, particularly in the figure without a subject end term, to be misled into erroneously treating *all*-premisses as founding premisses. As the term-ordering data attests, there is a strong tendency to avoid beginning a description with negative information (presumably for the obvious information theoretic reasons), and so it is not suprising that *all* exercises an influence that *none* does not. One of the virtues of appreciating the centrality of case identifiability to syllogistic reasoning is that it connects it with more general text comprehension. Building up specifications of individuals is a routine part of that process.

On the second problem, formulation of abstract conclusions, data from the new individuals task which lacks this component problem strongly suggests that the problem of finding quantifiers for conventional conclusions is a better predictor of difficulty than the number of possible models. When the right quantifier does not appear in either premiss, the problem is hard even in a one-model example. Describing data at the right level of abstraction is important for discovering regularities.

The second way forward is to seek more detailed evidence about processes occurring between input and output—evidence capable of revealing implementational detail. We can expect this to be a hard task. It is notoriously difficult even for mental processes much simpler than syllogism solution, as the imagery debate attests. Some beginnings have been made using dual tasks, thinking aloud protocols, phenomenological correlations, and other methods.

Matsuno (1987) is one of the few experimenters who has collected 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. We do not doubt that different implementations are at work, or that these bear some relation to the phenomenology of imagistic or sentential representation. We just doubt the efficacy of available evidence for discriminating them.

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.

What can the mapping between graphical and sentential methods presented here tell us about where to look for the difference between visual and verbal thinking? One might reasonably suppose that the computational differences between external diagrams and external language should be our best guide to understanding internal mental processes. Although the particular graphical and sentential systems presented here applied to the same syllogistic domain can be shown to correspond, there are important general differences between graphical and sentential semantics. Stenning, Inder & Neilson 1995 review some of these differences.

One of the most relevant differences for the subject first encountering syllogistic reasoning is evidence of semantic symmetry. All the sentences concerned are asymmetrical—quantifier precedes subject precedes predicate. It is the nature of sentences that they are asymmetrical—they have a beginning and an end. But the semantically symmetrical quantifiers are *graphically* symmetrical, and the semantically asymmetrical quantifiers are *graphically* asymmetrical. The diagrams for *some* and *none* are symmetrical. The diagrams for *all* and *some ... not* are asymmetrical (note the importance of the $\times$ notation in establishing this correspondence).

There is evidence that grasping the conditions under which sentences are convertible ($QuantifierAB \equiv QuantifierBA$) in this new kind of discourse is one of the major problems for substantial numbers of students. This is one of the basic tenets of conversion theories. Stenning, Cox & Oberlander (1995) and Stenning, Yule & Cox (1996) show that there are a small number of distinctive patterns of quantifier interpretation in this population and that these patterns are best described in terms of attitudes to the issue of convertability—or semantic symmetry.

Learning the nature of pure deductive discourse in the syllogism poses the problem of unlearning functions carried by the subject/predicate distinction in natural language and expository discourse (Stenning in press). Here is a purchase for computational differences between graphics and sentences to play a part in the learning processes involved.

Once the problem is seen as finding semantic and computational characterisations of 'visual' and 'verbal' refinements of a core abstract process, the search for the *right* system of mental representation becomes far less compelling. Once it seems highly plausible that different people, or the same person at different stages of expertise, use different kinds of representation, there is much less temptation to organise psychological research into the search for the holy grail of the one true mental representational system. Instead a picture emerges of people searching for, or constructing, representational systems suitable to the tasks they encounter.

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 represention al 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 of why this might be so.

5 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. We would further like to thank the three anonymous referees whose comments helped greatly with improving our exposition.

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.

6 References

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

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**, 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.

Lemmon, E.J. (1965) *Beginning Logic.* Walton-on-Thames: Nelson.

Levesque, H. (1988) Logic and the complexity of reasoning. *Journal of Philosophical Logic,* **17** 355–389.

Marr, D. (1982) *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: W.H. Freeman.

Matsuno, T. (1987). Cognitive style and representational strategies in categorial syllogistsic reasoning. *Tohoku Psychologica Folia,* **46**, 97–102.

Pylyshyn, Z. (1973). What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imagery. *Psych Bulletin,* **80**, 1–24.

Schubert, L. (1976) Extending the expressive power of semantic networks. *Artificial Intelligence*, **7(2)**, 163–198.

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*, 282–306. Montgomery, Vt.: Bradford, 1981.

Stenning, K. (1996) Embedding logic in communication: lessons from the logic classroom. In *Logic and Argumentation*. Proceedings of the Colloquium on Logic and Argumentation, Royal Netherlands Academy of Arts and Sciences, Amsterdam, June 1994. Pps. 227–240. Amsterdam: North Holland.

Stenning, K., Inder, R. & Neilson, I. (1995). Applying semantic concepts to the media assignment problem in multi-media communication. In *Diagrammatic Reasoning: Computational and Cognitive Perspectives on Problem Solving with Diagrams*. Glasgow, J., Narayanan, N. & Chandrasekaran, B. (eds.) MIT Press for AAAI Press: Cambridge. pps. 303–338.

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. 446–486.

Stenning, K. & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: logic and implementation. *Cognitive Science,* **19**, 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)**, 333–354.

Stenning, K., Cox, R. & Oberlander, J. (1995). Attitudes to logical independence: traits in quantifier interpretation. Proceedings of the Seventeenth Meeting of the Cognitive Science Society, Pittsburgh 1995. pps. 742–747.

Stenning, K. Yule, P. & Cox, R. (1996). Quantifier interpretation and syllogistic reasoning. Proceedings of the Eighteenth Meeting of the Cognitive Science Society, UCSD, La Jolla Ca. July. pps. 201–206.

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.

Yule, P. (1996a). *Logic and Implementation in Human Reasoning: the Psychology of Syllogisms.* PhD Thesis, University of Edinburgh.

Yule, P. (1996b) A `prolog` implementation of the Method of Euler Circles for syllogistic reasoning. H.C.R.C. Technical Reports Series (HCRC/TR-78). Human Communication Research Centre, University of Edinburgh. Available online at
`http://www.cogsci.ed.ac.uk/hcrc/publications/tr-78.ps.gz`.

A Appendix: Logic and implementation details

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

Second Premiss	First Premiss Aab	Iab	Eab	Oab	Aba	Iba	Eba	Oba
Abc	abc 1	abc 1	-abc		abc	abc 1	-abc	-abc 1
Ibc			-abc 2		abc 2		-abc 2	
Ebc	ab-c 1 -a-bc 2	ab-c 1	-ab-c		ab-c	ab-c 1	-ab-c	-ab-c 1
Obc			-ab-c 2		ab-c 2		-ab-c 2	
Acb			a-b-c 1 -abc 2	a-b-c 1	abc 2		a-b-c 1 -abc 2	
Icb			-abc 2		abc 2		-abc 2	
Ecb	ab-c 1 -a-bc 2	ab-c 1	-ab-c		ab-c	ab-c 1	-ab-c	-ab-c 1
Ocb	-a-bc 2							

Table 13: Valid individual conclusions for each premiss pair. Numerals denote source premiss number when unique.

A.2 Problem Types and Registration Diagrams

Problem types are sets of logically equivalent problems, abstracting over premiss order (and hence end-term assignment) and over the valid conversion of I and E premisses. Each problem type corresponds to a different registration diagram, since the graphical system naturally expresses these abstractions (see Section 2.2.1). Figures 9 and 10 show the full set of problem types, for problems with and without valid conclusions respectively. Each row of each table gives, in order, the name[11] of the problem type, the corresponding registration diagram, and the full set of instance problems using the standard mnemonics (A = all, I = some, E = no, O = some...not), broken down by figure.

A.3 Sentential representations

A.4 Glossary of logical terms

B Appendix: Experimental details

B.1 Experimental materials: vocabulary

B.2 Raw data tables

[11]The names have some mnemonic significance: the first letter specifies whether the problems have valid conclusions (v) or not (n); the second letter indicates whether both premisses are positive (p), at least one is negative (n) or the problem type has U-conclusions (u); the final digit just reflects the listing order.

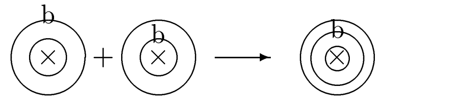
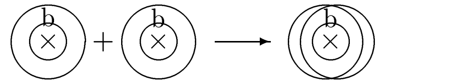
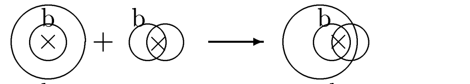
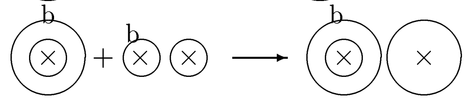
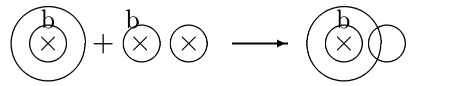
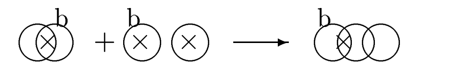
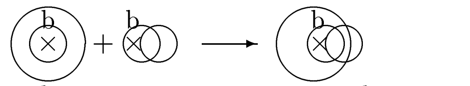
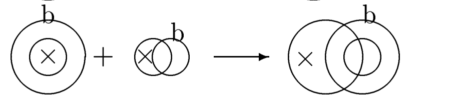
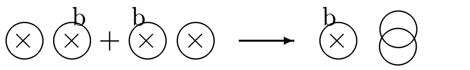
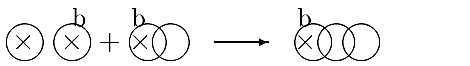
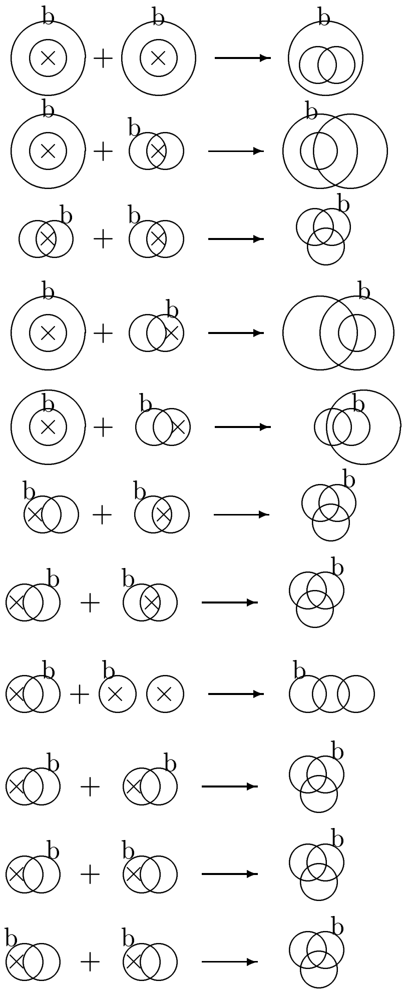
Problem Type	Registration Diagram	ab/bc	ba/cb	ab/cb	ba/bc
		Problem instances			
vp1		**A**A	A**A**		
vp2					AA
vp3		**I**A	A**I**		**I**A A**I**
vn1		**AE**	**EA**	**AE** **EA**	
vn2		EA	AE		AE EA
vn3		**I**E E**I**	**I**E E**I**	**I**E E**I**	**I**E E**I**
vn4					**O**A A**O**
vn5				**O**A A**O**	
vu1		EE	EE	EE	EE
vu2		E**O**	**O**E		**O**E E**O**

Figure 9: The set of problem types *with* valid conclusions, showing registration diagrams and instance problems. Unique source premisses are indicated by **bold**.



Problem Type	Problem instances ab/bc	ba/cb	ab/cb	ba/bc
np1			AA	
np2	AI	IA	IA AI	
np3	II	II	II	II
nn1	AO	OA		
nn2	OA	AO		
nn3	IO	OI		IO OI
nn4	OI	IO	IO OI	
nn5	OE	EO	OE EO	
nn6			OO	
nn7	OO	OO		
nn8				OO

Figure 10: The set of problem types *without* valid conclusions, showing registration diagrams and instance problems.

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

Table 14: Sentential representations of valid syllogisms with discourse introducing property (in **bold**) and use of modus tollens (MT)

Term	Definition
consequence relation	The function which maps sets of true premisses onto true conclusions, thus defining what arguments are valid in a logic. The most abstract characterisation of a logic.
implementation	Provision of a mechanism that will compute an abstract function (such as finding a conclusion from premisses supplied)
model-theoretic	That part of logic which deals with consequence relations in terms of the sets of things which are models of their premisses and conclusions. In general, an argument is valid just in case the conclusion is true in *all* models of its premisses
case-identifiability	Metalogical property of the syllogism whereby any pair of premisses which have a valid conclusion entail the existence of a type of individual specified in terms of all three properties.
U-conclusion	A conclusion of the form *Some* $\neg a$ *is* $\neg b$ which is not provided in the classical syllogism.
quantifier	In the syllogism, *all, some, no, some ... not.*
universal	One of the two quantifiers *all* and *no*
particular	Traditional term for *existential*
existential	One of the two quantifiers *some* or *some ... not*
end term	Term occurring in only one of the premisses of a syllogism, here represented as a or as c
middle term	Term occurring in both premisses of a syllogism, here represented as b
figure	One of the four grammatical arrangements of terms in a syllogism
figural effects	effects of figure on the frequencies subject/predicate structures of conclusions (e.g. *Some a are c* vs. *Some c are a*).
mood	Traditional term for quantifier
validity	An argument is valid if, whenever its premisses are true, then its conclusion is also true
type	A kind of individual satisfying a description
model (of sentence set)	An assignment from a vocabulary of terms onto a set of things (in the case of the syllogisms *types*) which make all the members of the set of sentences true
maximal model	The model of a syllogism containing the most possible types
minimal model	A model of a syllogism containing the least possible types
minimal region	A closed curve in an Euler diagram not bisected by any line

Table 15: Some logical terms and their definitions (continued below)

Term	Definition
critical individual	A type of individual required to exist by a syllogism.
source premiss	The premiss which provides the clause matching either the antecedent (for modus ponens) or the consequent (for modus tollens) of the conditional premiss
conditional premiss	The universal premiss which functions as a conditional
monadic predicate logic	Predicate logic analyses sentences into quantifiers, predicates or relations, and terms. Monadic predicate logic has only one-place predicates but no multi-place relations
propositional logic	Propositional logic analyses the relations between unanalysed propositions marked by connectives such as *if ... then, not, and, or...*
rule of inference	A rule of inference requires premisses matching its left hand side, and rewrites them as premisses matching its right hand side. Each connective requires an inference rule for its introduction into formulae and one for its elimination from formulae.
&-elimination	An inference rule for eliminating & from formulae.
&-introduction	An inference rule for introducing & to formulae.
$\rightarrow -elimination$	An inference rule for eliminating the conditional connective—classically known as *modus ponens*
modus ponens	see $\rightarrow -elimination$
modus tollens	The inference rule which takes as premisses a conditional and the negation of its consequent and returns the negation of its antecedent
conditional	A sentence of the form *If ... then ...*, of which the *if* clause is the *antecedent* and the *then* clause is the consequent. In propositional logic, universal quantifiers take on the form of conditionals. Hence each syllogism with a valid conclusion has one premiss functioning as a conditional
antecedent	The *if* clause of a conditional
consequent	The *then* clause of a conditional
self-consistent representation system	One in which it is not possible to represent inconsistent sets of premisses e.g. the Euler system presented here cannot represent *All a are b* and simultaneously *Some a are not b*

Table 16: Some logical terms and their definitions (continued)

Profession	*Nationality*	*Interest*
butchers	Greeks	cricketers
musicians	Italians	chessplayers
archaeologists	Irish	boxers
waiters	Swiss	houseowners
politicians	Australians	archers
judges	Hawaiians	beerdrinkers
joiners	Kenyans	hillwalkers
managers	Hungarians	vegetarians
vets	English	footballers
architects	Koreans	skiers
artists	Canadians	squashplayers
students	Indonesians	horseriders
doctors	Israelis	hangliders
dentists	Filipinos	pessimists
grocers	Dutch	dieters
cooks	Egyptians	golfers
counsellors	Turks	gourmets
nurses	Japanese	christians
plumbers	Spanish	atheists
lawyers	Ethiopians	liberals
bakers	Tibetans	tennisplayers
miners	Sudanese	cyclists
typists	Indians	gymnasts
stockbrokers	Russians	rowers
domestics	Poles	dogowners
sailors	Chinese	shotputters
teachers	Swedes	swimmers
physicists	Brazilians	highjumpers
builders	Germans	iceskaters
busdrivers	French	layabouts
clerks	Mexicans	moslems
chemists	Americans	pacifists

Table 17: Vocabulary used in both experiments.

Type	Prob.	Conclusions	Valid %
vp1	AA1	abc:21, acb:1	100.0
	AA2	cba:9, bac:8, abc:1, acb:1, bca:1, *NVC:1*	90.9
vp2	AA4	bac:15, bca:5, acb:1, *NVC:1*	95.5
vp3	IA1	abc:17, bca:4, *NVC:1*	95.5
	AI2	cba:13, bac:5, acb:1, bca:1, *NVC:2*	90.9
	IA4	bca:10, bac:9, cab:1, cba:1, *NVC:1*	95.5
	AI4	bac:9, bca:8, abc:3, *ab-c:1, NVC:1*	90.9
vn1	AE1	ab-c:19, ba-c:1, b-ca:1	95.5
	EA2	cb-a:17, bc-a:1, -ba-c:1, *cba:3*	86.4
	AE3	ab-c:12, -cba:1, c-b-a:4, c-a-b:2, *ac-b:1, NVC:2*	86.4
	EA3	cb-a:12, bc-a:2, -abc:1, b-ac:1, a-b-c:3, *NVC:2*	86.4
vn2	EA1	bc-a:11, -abc:2, b-ac:2, cb-a:1, *ca-b:1, a-b-c:1, NVC:3*	72.7
	AE2	ba-c:12, ab-c:1, -cba:1, *c-b-a:1, c-a-b:2, NVC:5*	63.6
	EA4	bc-a:17, cb-a:2, b-ac:1, *-bac:1, NVC:1*	90.9
	AE4	ba-c:18, ab-c:2, *ac-b:1, bc-a:1*	90.9
vn3	EI1	bc-a:12, cb-a:2, -abc:3, *ca-b:1, bca:1, NVC:3*	77.3
	IE1	ab-c:17, *ac-b:1, c-b-a:1, NVC:3*	77.3
	EI2	cb-a:15, *-cb-a:1, b-a-c:1, cba:1, NVC:2*	68.2
	IE2	ba-c:12, ab-c:1, *ac-b:1, NVC:8*	59.1
	EI3	cb-a:14, -abc:1, *-ab-c:1, -cb-a:1, NVC:3*	68.2
	IE3	ab-c:7, *NVC:4*	31.8*
	EI4	bc-a:13, b-ac:2, *bac:1, -ba-c:1, a-b-c:1, -cb-a:1, NVC:2*	68.2
	IE4	ba-c:13, b-ca:3, ab-c:3, *NVC:3*	86.4
vn4	AO4	ba-c:14, ab-c:2, b-ca:2, *cba:1, NVC:3*	72.7
	OA4	bc-a:15, b-ac:3, cb-a:1, *bac:1, NVC:2*	86.4
vn5	AO3	c-b-a:9, c-a-b:1, -a-bc:1, *acb:1, ab-c:1, NVC:8*	50.0
	OA3	a-b-c:5, -c-ba:2, *abc:1,cb-a:2,b-ac:1,bc-a:1,-a-c-b:1,-c-b-a:1, NVC:7*	31.8
vu1	EE1	b-c-a:6, b-a-c:4, -ab-c:1, *ca-b:1, c-b-a:1, -c-b-a:1, NVC:6*	50.0
	EE2	b-a-c:6, b-c-a:1, -cb-a:1, *a-b-c:1, ca-b:1, NVC:10*	36.4
	EE3	b-a-c:8, -ab-c:1, b-c-a:1, *ac-b:1, c-b-a:1, -a-b-c:1, NVC:9*	45.5
	EE4	b-a-c:12, -ab-c:1, -a-cb:1, *-bac:1, ca-b:1, -b-a-c:1, NVC:4*	63.6
vu2	EO1	b-c-a:10, -ab-c:1, -a-cb:1, b-a-c:1, *a-b-c:1, NVC:7*	59.1
	OE2	b-a-c:11, b-c-a:2, *NVC:8*	59.1
	EO4	b-c-a:9, b-a-c:5, *b-ac:2, ab-c:1, b-a-c:1, NVC:4*	63.6
	OE4	b-a-c:10, b-c-a:5, -c-ab:1, *ba-c:2, ca-b:1, NVC:3*	72.7

Table 18: Individuals Task raw data for all problems with valid conclusions. Invalid responses are given in *italics*.

Type	Prob.	Invalid Conclusions	NVC	NVC %
np1	AA3	*abc:4, cab:3, bac:2, -b-a-c:2*	11	50.0
np2	AI1	*abc:10, cab:2, bac:1*	9	40.9
	IA2	*acb:4, cba:4, bac:2, bca:1*	11	50.0
	AI3	*abc:2, cba:1, acb:1, cab:1, bac:1*	16	72.7
	IA3	*abc:2, cba:1, bac:1, cab:1*	15	68.2
np3	II1	*abc:3, bac:1*	18	81.8
	II2	*abc:1, cba:1, bac:1, bca:1*	18	81.8
	II3	*abc:1, cab:1, bac:1*	18	81.8
	II4	*bac:4, cba:1, cab:1*	16	72.7
nn1	AO1	*ab-c:8, ba-c:2, a-b-c:1*	11	50.0
	OA2	*cb-a:11, acb:1*	9	40.9
nn2	OA1	*-abc:1, cb-a:1, abc:1, bc-a:1, a-bc:1, ca-b:1, a-b-c:2*	12	54.5
	AO2	*ba-c:3, b-ca:1, ab-c:1, c-b-a:1, c-a-b:1*	15	68.2
nn3	IO1	*ab-c:4, abc:2*	16	72.7
	OI2	*cb-a:3, b-ac:2, bca:1*	16	72.7
	OI4	*bc-a:2, b-ac:1, cb-a:1*	18	81.8
	IO4	*ba-c:2, b-ac:1, bac:1*	18	81.8
nn4	OI1	*bac:1, bc-a:1, a-bc:1*	18	81.8
	IO2	*abc:1, bac:1, ab-c:1*	19	86.4
	OI3	*abc:1, cb-a:1*	19	86.4
	IO3	*ab-c:1, cba:1, ba-c:1*	19	86.4
nn5	OE1	*a-bc:1, ca-b:1, ab-c:1, b-ca:1, b-a-c:1, c-a-b:1*	16	72.7
	EO2	*b-a-c:2, -bac:1, ca-b:1, b-c-a:1*	17	77.3
	OE3	*a-b-c:3, ab-c:2*	16	72.7
	EO3	*-ab-c:1, bc-a:1, b-a-c:1*	19	86.4
nn6	OO3	*a-bc:1, ca-b:1, b-c-a:1*	19	86.4
nn7	OO1	*b-a-c:1, ca-b:1*	20	90.9
	OO2	*b-a-c:1, bc-a:1, ca-b:1*	19	86.4
nn8	OO4	*bac:1, b-a-c:1*	20	90.9

Table 19: Individuals Task raw data for all problems without valid conclusions. Invalid responses are given in *italics*.

Type	Problem	Aac	Iac	Eac	Oac	Aca	Ica	Eca	Oca	NVC	Valid %
vp1	AA1	18	4							*3*	88
	AA2	*6*	1			17				*1*	72
vp2	AA4	*6*	8			*4*	5			*2*	52
vp3	IA1	*1*	21				3			*0*	96
	AI2		1				21		*2*	*1*	88
	IA4		12		*1*		12			*0*	96
	AI4		16				8		*1*	*0*	96
vn1	AE1			19		*1*		3		*1*	88
	EA2			6				18		*0*	96
	EA3			9				12		*4*	84
	AE3			16				7		*2*	92
vn2	EA1	*1*	*1*	*7*			*1*	*3*	4	*8*	16
	AE2			*4*	3			*3*		*15*	12
	EA4			*7*				*6*	7	*5*	28
	AE4			*8*	9			*2*		*6*	36
vn3	EI1			*2*	*5*				5	*11*	20
	IE1				17				*1*	*7*	68
	EI2			*1*			*1*		15	*6*	60
	IE2		*2*		5			*1*	*4*	*12*	20
	EI3			*3*	*3*				9	*9*	36
	IE3		*1*		8			*1*		*15*	32
	EI4				*1*		*1*	*1*	14	*8*	56
	IE4			*1*	11					*12*	44
vn4	OA4		*6*		*2*		*4*		8	*4*	32
	AO4		*2*		14		*1*		*2*	*4*	56
vn5	OA3		*3*		10				*1*	*10*	40
	AO3				*3*		*3*		5	*11*	20

Table 20: Standard Task raw data for all problems with valid conclusions. Invalid responses are given in *italics*.

Type	Problem	Aac	Iac	Eac	Oac	Aca	Ica	Eca	Oca	NVC	NVC %
vu1	EE1			*2*						23	92
	EE2		*1*					*1*		23	92
	EE3			*2*			*1*	*2*		20	80
	EE4			*1*	*1*			*1*		22	88
vu2	EO1		*1*		*1*				*3*	19	76
	OE2				*2*				*2*	21	84
	OE4				*11*				*3*	10	50
	EO4				*2*				*7*	14	56
np1	AA3	*5*				*3*	*1*			16	64
np2	AI1		*5*				*4*		*2*	14	56
	IA2		*3*			*1*	*5*		*1*	15	60
	IA3		*6*				*4*			13	52
	AI3		*2*		*1*		*5*		*1*	15	60
np3	II1		*4*		*1*		*1*		*1*	18	72
	II2						*2*		*1*	22	88
	II3		*3*		*1*		*1*		*1*	18	72
	II4		*2*		*2*		*1*			20	80
nn1	AO1		*1*		*7*				*1*	15	60
	OA2		*1*		*1*		*2*		*11*	9	36
nn2	OA1		*5*		*6*		*1*		*3*	8	32
	AO2		*1*		*1*		*8*		*9*	4	16
nn3	IO1				*7*	*1*	*1*			16	64
	OI2				*1*		*1*		*4*	18	72
	OI4				*1*				*3*	21	84
	IO4				*2*				*2*	21	84
nn4	OI1		*2*		*3*				*1*	18	72
	IO2				*2*		*3*			20	80
	OI3				*1*				*5*	19	76
	IO3						*1*		*3*	20	80
nn5	OE1				*10*				*1*	13	52
	EO2		*1*		*2*		*3*		*5*	13	52
	OE3				*2*			*1*	*2*	20	80
	EO3		*1*		*1*		*1*		*9*	13	52
nn6	OO3				*1*		*1*			23	92
nn7	OO1		*1*		*3*					21	84
	OO2				*1*				*3*	21	84
nn8	OO4				*1*					24	96

Table 21: Standard Task raw data for all problems without valid conclusions. Invalid responses are given in *italics*.