This is an accepted manuscript (6 January 2016) of an article to be published by Wiley in *Cognitive Science,* available on line:
http://onlinelibrary.wiley.com/journal/10.1111/%28ISSN%291551-6709/earlyview

Preadolescents Solve Natural Syllogisms Proficiently

Guy Politzer[1]

Christelle Bosc-Miné[2]

Emmanuel Sander[2]

1. Institut Jean Nicod, Ecole normale supérieure, 29 rue d'Ulm, 75005 Paris, France. (politzer@univ-paris8.fr)

2. Christelle Bosc-Miné (christelle.bosc@u-picardie.fr) and Emmanuel Sander (emmanuel.sander@univ-paris8.fr), University of Paris-8, Department of Psychology, laboratory Paragraphe, CRAC, 2 rue de la Liberté, 93528 Saint-Denis, France.

Keywords: reasoning - categorical syllogisms - ecthesis - categorization - pragmatics - argumentation - preadolescence

Abstract

"Natural syllogisms" are arguments formally identifiable with categorical syllogisms that have an implicit universal affirmative premise retrieved from semantic memory rather than explicitly stated. Previous studies with adult participants (Politzer, 2011) have shown that the rate of success is remarkably high. Because their resolution requires only the use of a simple strategy (known as *ecthesis* in classic logic) and an operational use of the concept of inclusion (the recognition that an element that belongs to a subset must belong to the set but not vice versa), it was hypothesised that these syllogisms would be within the grasp of non-adult participants, provided they have acquired the notion of deductive validity. Here, 11-year-old children were presented with natural syllogisms embedded in short dialogues. The first experiment showed that their performance was equivalent to adults' highest level of performance in standard experiments on syllogisms. The second experiment, while confirming children's proficiency in solving natural syllogisms, showed that they outperformed children who solved non-natural matched syllogisms in the same experimental setting. The results are also in agreement with the argumentation theory of reasoning.

1. Introduction

For decades, the psychological study of syllogistic reasoning has established itself as a classical paradigm in which the logical forms of the syllogisms are instantiated in natural language. Typically, participants are asked to imagine a situation in which various relations hold between objects or individuals. These relations are expressed by quantified statements that relate two terms in each of the two premises (a brief reminder of the structure and denomination of syllogisms can be found in Appendix A). The relations between the terms of the premises are arbitrarily designed for, and attached to, the fictitious situation or scenario under consideration. For instance, participants are told that at a horticultural competition, *all the roses are hybrids;* and *all the roses are prizewinners*; then they may be presented with sentences that relate *prizewinners* and *hybrids* quantified by *all, no, some, some...not,* to decide whether any one necessarily follows from the premises, or they may be asked to produce their own conclusion. Here a conclusion necessarily follows. Readers unfamiliar with syllogisms may wish to have a try at deriving the conclusion before going further. The majority of people answer *all the prizewinners are hybrids,* which is incorrect. Indeed, there may be prizewinners (other than roses) that need not be hybrids. The correct conclusion is *some prizewinners are hybrids;* (*some hybrids are prizewinners* is a logically equivalent conclusion). There is some artificiality in the task, and this is even more acute in abstract variants in which no situation is described and the terms are just letters or symbols.

After more than a century of research based on this paradigm, the mechanisms which people use in the laboratory to attain a solution are yet unknown —at least no consensus has emerged. In the most recent and comprehensive review, Khemlani and Johnson-Laird (2012) write that no extant theory provides an adequate account of the data. Based on the meta-analysis which they present, it appears that around one half of the syllogisms are solved less than 50% of the time, which we regard as an indication of poor overall performance. In addition, the main upshot of the studies is that there is extreme variability in performance as a function of logical form, to such an extent that a few syllogisms are seldom solved, while a few others are solved most of the time. The fact that performance is poor and that the solution of many logical forms is out of the reach of a majority of individuals (and of most of them in a few cases) stands in sharp contrast to reasoners' proficiency in producing and understanding arguments in daily life. One cannot but question the adequacy of the traditional syllogistic task if it is used to study such arguments. Still, we do not claim that lay people do not or cannot engage in syllogistic arguments —this paper will be devoted to show quite the opposite. We believe that the investigation of syllogistic reasoning requires a different task inspired by people's arguments in natural settings. In what follows, we will pursue the development of a new syllogistic task based on the notion of "natural syllogism", initially described in Politzer (2011).

Natural syllogisms are categorical syllogisms occurring in various guises in ordinary conversation which are defined by two joint properties. One is formal: they are classical syllogisms whose minor premise is a universal affirmative sentence (an A sentence). The other is semantic: the minor premise contains an inclusion relation between two categories stored in long-term memory. This latter property has the pragmatic consequence that natural syllogisms are typically enthymematic, that is, their minor premise need not be stated explicitly (hence the superficial difference with the formal description). The semantic and pragmatic characteristics distinguish them from their counterparts used in the traditional laboratory task. There are other non essential features, such as the fact that, like in other naturally occurring arguments, the place of the conclusion is variable: it may come first and be subsequently justified by the premises, or it may be presented last [note 1]. Let us compare an artificial and a natural syllogism in the same mood and figure. The artificial example given above is repeated here:

all the roses are hybrids; all the roses are prizewinners;

therefore: *some prizewinners are hybrids.*

Its natural counterpart is:

all the roses are hybrids;

therefore: *some flowers are hybrids;*

(implicit premise: *all roses are flowers).*

Both are instantiations of the AA-3 syllogism *all M are P; all M are S* whose correct conclusion is *some S are P.* They have in common the M term (roses), the P term (hybrids) but a different S term (*prizewinners* and *flowers*, respectively). So, the difference between the two instantiations lies in the minor premise, which is a contingent relation in the former case (*all the roses are prizewinners)* but a knowledge-based inclusion relation between categories in the latter case (*all roses are flowers)*.

Readers can appreciate how easy it is to get the conclusion in the natural case as compared to the artificial case. The overall ease of solving natural syllogism has been confirmed by experimental data with adult participants reviewed below. The aim of the present paper is to offer more evidence of lay reasoners' competence in natural syllogistic reasoning and to detail the theoretical explanation outlined in Politzer (2004, 2011). This explanation is based on the assumption that *the inclusion relation in the minor premise has the power to prime an effective method of derivation of the conclusion,* called "method of exposition" by medieval logicians, and "ecthesis" in Aristotle's writings (Kneale & Kneale, 1962) which we now describe.

2. The Method of Proof by Ecthesis

2.1. Ecthesis applied to classical syllogisms

2.1.1. Contemporaneous views on the syllogistic and ecthesis

Although Aristotle's syllogistic had been the principal reference for two millennia, its supremacy did not survive the developments of logic in the 19th century and more critically in the early 20th century. However, there has been a revival of interest for it in the wake of the work of Lukasiewicz (1957) who treated the syllogistic as an autonomous system. Various formalizations have been proposed (e. g., Corcoran, 1974; Patzig, 1968; Rose, 1968; Smiley, 1973) using fragments of predicate logic. Interestingly, and importantly for our purposes, a number of logicians have also analyzed the concept of ecthesis from a modern point of view. They generally agree to regard it as a precursor of the rule of existential instantiation (Lear, 1980; Mignucci, 1991; Smiley, 1973; Smith, 1982; Thom, 1981). Thom (1981) presents a version of the syllogistic that incorporates ecthesis among its rules of inference which can be viewed as providing a computational model of syllogistic reasoning. In sum, ecthesis applied to Aristotelian syllogisms is grounded in centuries of reflection and accommodated in the technical tools of modern predicate logic. Therefore, it suggests itself as a theoretical framework, logically and psychologically.

2.1.2. The principle of the method

Solving a syllogism amounts to answering the following question: Can an *individual* object identified by two attributes, namely the middle term M and one of the end terms (say S) be characterized by a third attribute P (or not-P)? To do so, one "exposes" (i. e., selects) an individual in a premise; it is always doubly predicated because the affirmative quantified sentences *(all/some S are M, all/some M are S)* allow the selection of an individual that is both S and M, and similarly the two negative sentences *(no S are M, some S are not M)* allow the selection of an individual that is both S and not-M. Then, using the other premise, one searches for a triple predication in terms of M, S, P (or their negation) of this individual. If the search fails, that is, if no such triple predication exists, the syllogism is invalid. If the search succeeds in showing that there exists at least one such individual, the syllogism is valid. The M term is deleted, leaving an individual doubly predicated in terms of S and P (or its negation), which constitutes the core of the conclusion. There remains to determine whether the conclusion is particular or universal, which will be detailed shortly.

2.1.3. The complete ecthetic procedure

We assume that the mental representation of the four sentences is based on generalized quantification (Politzer, Van der Henst, Delle Luche & Noveck, 2006): *all P are Q* and *some P are not Q* are represented as the inclusion and the non inclusion of P in Q, respectively; and *no P are Q* and *some P are Q* are represented as an empty and a non empty intersection of P and Q, respectively. We now examine the ecthetic procedure in some detail, applying it first to the artificial case, and showing that it may be more or less sophisticated depending on the logical form of the syllogism. We present the formal expression together with its artificial instantiation and begin with the simplest case, the AA-1 syllogism:

all M (prizewinners) are P (hybrids)

all S (roses) are M (prizewinners)

Select an individual S (rose), which the second premise (the minor premise) warrants to be an M (prizewinner), and try to apply to this SM (rose prizewinner) individual the predication expressed by the first premise (the major premise): because this SM (rose prizewinner) is an M (prizewinner), it is warranted to be a P (hybrid), that is, this SM is warranted to be a P *qua* M), hence the assurance that there exists an SMP (rose prizewinner hybrid) and therefore *some S (rose) is a P (hybrid).* We could stop at this conclusion but the exposition of the individual S (rose) can be repeated for any rose (and such a repetition would exhaust the set of roses), so that the final conclusion is the universal sentence *all S (roses) are P (hybrids).* Because we always start with a single individual and conclude with respect to some or all the individuals under consideration, this search for generalization that will result in a particular or a universal conclusion concerns all the syllogisms. It has a role analog to a subroutine that will be called <EXHAUST>. Its result depends on the quantity of the major premise (a particular sentence cannot produce a universal sentence) and also, less trivially, on the direction of the set relation in the minor premise. In the present case the universal conclusion stems from the fact that attributing P to every SM implies attributing P to every S. This point is further illustrated by the opposite result that is obtained for the AA-3 syllogism (presented earlier and repeated here):

all M (roses) are P (hybrids)

all M (roses) are S (prizewinners)

Select an individual M (rose), which is warranted to be an S (prizewinner); this MS (rose prizewinner) is also warranted to be a P (hybrid), hence the triple predication MSP (rose prizewinner hybrid) and like above the assurance that *some S (prizewinner) is a P (hybrid).* However, applying < EXHAUST >, we must ask whether attributing P to every SM implies attributing P to every S, and the answer is negative: indeed, we cannot go further because the MS is waranted to be a P *qua* M, but not *qua* S, so that there is no warrant that when exposing the M (roses) repeatedly all the S (prizewinners) will be selected and so the conclusion is particular.

This procedure enables one to solve EA-1 (which has a universal conclusion) and all the syllogisms in the third figure (which have a particular conclusion). The other two syllogisms in the first figure are IA-1 (some M P, all S M) and OA-1 (some M not P, all S M). For IA-1:

some M (prizewinners) are P (hybrids)

all S (roses) are M (prizewinners)

After an SM (rose prizewinner) individual has been selected there is no possibility to find a triple predication because there is no guarantee of co-reference between this SM (rose prizewinner) and any of those specific M (prizewinners) that are P (hybrids), that is, an SM (rose prizewinner) may be an MP (prizewinner hybrid) but it may be an M not-P (prizewinner not-hybrid). In addition, if we try exposition from the major premise, after selecting an MP (prizewinner hybrid) individual, the minor premise does not offer information about the possibility of predicating it with S (rose). Consequently there is no conclusion, the syllogism is invalid. The same result is obtained for OA-1 after a similar check for co-reference, acting again as the analog of a subroutine which we call the <CO-REF>. It is triggered whenever the premise that provides the third attribute is a particular sentence.

We consider now syllogisms in the fourth figure. The simple procedure using <EXHAUST> applies to AA-4 (all P M, all M S) and IA-4 (some P M, all M S). It is slightly more difficult for EA-4 (no P M, all M S) and OA-4 (some P not M, all M S).

For EA-4:

no P (prizewinners) is M (roses)

all M (roses) are S (hybrids)

Select an MS (rose hybrid) and try a triple predication using in the major premise an individual that is P (prizewinner): adding hypothetically the attribute P results in a contradiction. Indeed, an MS (rose hybrid) cannot be qualified by P (prizewinner) because no P (prizewinner) is M (rose), which warrants that *some S (hybrid) is not a P (prizewinner).* We have used a third subroutine, <SUPPOSE>, which here ends with a form of *reductio ad absurdum*.

There is a subtle difference with the OA-4 (some P not M, all M S) pair:

some P (prizewinners) are not M (roses)

all M (roses) are S (hybrids)

where <SUPPOSE> aborts. Consider as above an MS (rose hybrid) and suppose it to be a P (prizewinner): we will not know whether or not this individual is one of those that are not M (rose) so that no decision follows. <SUPPOSE> is blocked by <CO-REF> and does not yield. any decision. In addition, the identification starting from the major premise fails (and this applies to EA-4 as well) because after selecting a P (prizewinner) that is not M (rose) we lose track of it in the minor premise which informs about the M (roses) but not about the not-M (not-roses). In brief, OA-4 is invalid.

We finish with the syllogisms in the second figure: *P M; S M.* They have the peculiarity that no triple predication can be found because there is no middle term in subject position, that is, it is not possible for an individual characterized by S and M (or P and M) to be identifiable with a P individual (or an S individual, respectively) unless this P (or this S) is known to be a not-M (that is, unless <SUPPOSE> results in a *reductio)*. So, EA-2 (no P M, all S M) and OA-2 (some P not M, all S M) which have P and not-M individuals are valid and solved like EA-4 (no P M, all M S), while AA-2 (all P M, all S M) and IA-2 (some P M, all S M), which have P and M individuals, are invalid because when S is hypothetically attributed to P by <SUPPOSE>, it appears that this is not impossible (no contradiction ensues) but not necessary either, meaning that no conclusion follows.

Finally, it should be noticed that a few syllogisms can validly be changed into another one by converting a premise, which may lead to a simpler solution. This is the case of EA-2 (no P M, all S M) and EA-4 (no P M, all M S) which are more easily solved like EA-1 (no M P, all S M) and EA-3 (no M P, all M S), respectively, after the conversion of *no P are M* into *no M are P*.

There is no reason to expect ecthesis to be used spontaneously by reasoners in the traditional syllogistic task, except by skilled reasoners. Indeed, if it were, performance on this task would be high because all the syllogisms can be solved by this method (Thom, 1981), which we have just illustrated with a subset of 16 pairs [notes 2 and 3].

Note that several theoretical approaches posit a treatment which partly coincides with ecthesis. This is most clearly the case of Störring's (1908) historical study which describes a process of "insertion" that consists of selecting the end term of one premise and inserting it next to the middle term in the other premise. Similarly Braine (1998) assumed that good reasoners are able to apply a strategy he called the "choice of a secondary topic" which consists of selecting a doubly predicated subset (e. g., from *all M are S,* select *the S that are M).* In the same vein, Ford (1995) described a "substitution behaviour" which consists of replacing one term in a premise with another, like a substitution of variable in algebra, a variant of which, the "sophisticated" substitution, being necessary to solve the difficult syllogisms. The commonalities between these approaches are detailed in Politzer (2004). Finally, the theory proposed by Stenning and Oberlander (1995) and Stenning and Yule (1997) is entirely based on the notion of a triple characterization of individuals and can be regarded as a resuscitation of Aristotle's ecthesis. Within this approach a majority of participants (the "credulous" ones) interpret the premises in a way that leads to inappropriate choices of attributes. All this suggests that the method of exposition is too sophisticated for *most* reasoners to be discovered *but may be simple enough to be executed if a hint is offered to initiate it.* This hypothesis was tested by Politzer and Mercier (2008) who reasoned that simply suggesting the consideration of individuals to participants could prime ecthesis. To do this, they used singular syllogisms, that is, classical artificial syllogisms in which the quantifier "some" is replaced by a demonstrative (e. g., with x standing for an individual that belongs to a class X, *this x is a Y)* or by a singular determiner (e. g., *there is an x that is Y).* As predicted the rate of success increased, by about one third.

2.2. Ecthesis applied to natural syllogisms

We now turn to the resolution of natural syllogisms. It is immediately apparent that the special inclusion relation in their minor premise, namely an inclusion between categories that belong to a hierarchy stored in long term memory, suggests the use of ecthesis. To see this, consider again the AA-3 case (where the implicit minor premise is in braces):

all M (roses) are P (hybrids)

{all M (roses) are S (flowers)}

Upon hearing or reading the major premise *all the roses are hybrids,* the triple predication *rose-flower-hybrid* is triggered by the strength and availability of the rose-flower relation in a way that seems irrepressible [note 4]. This is the first and most important factor that applies to all the syllogisms under consideration.

There is a second factor: the three subroutines are readily available. Consider first <EXHAUST>, for instance, in the current AA-3 case: note that the triple predication *rose-flower-hybrid* is unlikely to lead to an *all* conclusion because it is obvious that the roses do not exhaust the flowers (all the roses constitute some flowers). Contrast with the artificial case in which the minor premise is instantiated by *all roses are prizewinners,* leading to a predication *rose-prizewinner-hybrid* and to the initial conclusion *some prizewinners are hybrids.* The application of <EXHAUST> which normally blocks the generalization to all the individuals, and to the erroneous response *all prizewinners are hybrids*, is far less available. Given *all roses are prizewinners,* the possibility that not all prizewinners are roses does not have the same saliency as knowledge that not all flowers are roses.

Similarly <CO-REF> is more available in the natural case. Compare the procedure given for IA-1 above with the natural case:

some M (flowers) are P (hybrids)

all S (roses) are M (flowers)

After exposing a rose, it is apparent that this need not co-refer with those flowers that are hybrids as they may not be roses.

Finally, <SUPPOSE> is likely to receive a short-cut in the natural case. Consider the OA-2 pair:

some P (prizewinners) are not M (flowers)

all S (roses) are M (flowers)

Can a P (prizewinner) not-M (not-flower) be an S (rose)? It is apparent that a non-flower cannot be a rose, so that a *reductio* is used implicitly without a need to explicitly develop it.

In summary, there exists a method to derive the solution of all categorical syllogisms, namely ecthesis, which in all likelihood most reasoners usually do not spontaneously use. However, for the natural syllogisms, which constitute a subset characterized by a minor premise referring to an inclusion relation between categories stored in semantic memory, ecthesis can be triggered by the inclusion relation and the use of its three basic procedures is readily available. Consequently, a good performance is expected. This prediction has already been supported by the results of a series of experiments (Politzer, 2011). Adult participants solved natural syllogisms in various formulations: the minor premise was stated explicitly, or primed by naming the hypernym of the inclusion relation, or left implicit. In all the cases, the rate of correct answers increased by about one half as compared to artificial control syllogisms. In the last condition the syllogisms were inserted in a dialogue and the mean rate of success was close to 80%.

The extension of these results to a younger population would be an important step to establish the generality of individuals' proficiency in natural syllogistic reasoning. Because participants should be able to distinguish between deductively valid and invalid syllogisms, there is a constraint on age. According to Byrnes and Overton (1986) discrimination between certain and uncertain conclusions (for conditional syllogisms) emerges in the fifth grade and the same obtains for formal deductive competence at large (Overton, 1990). Markovits, Schleifer and Mortier (1989) found that by 11 years of age children clearly respond differently to valid and matched invalid syllogisms. According to Moshman and Franks (1986) the concept of validity is still lacking at 9-10 years of age but typically develops between ages 10 and 12. Morris (2000) argues that this latter claim concerns explicit knowledge whereas children start developing implicit knowledge of validity as early as eight years. In brief, a population of preadolescents aged 10-11 seems appropriate to participate in a task in which they will be asked to evaluate the conclusion of deductive arguments such as natural syllogisms. Below, we present two experiments that aim to test the hypothesis that they master natural syllogisms. In the main experiment a group of children solving natural syllogisms will be compared with a control group solving the classic task. Before, we present the results of a pilot study subservient to the main experiment that aimed to finalize the method, especially the material and the analysis of verbal protocols.

3. Experiment 1

3.1. Method

3.1.1. Participants

Seventy-three children (38 girls) attending a primary school in two small French cities participated in the experiment. Their mean age was 11.0 (sd = 0.48).

3.1.2. Material and constitution of the syllogisms

In the present investigation we aim to assess the logical abilities that children exploit in daily life, especially when they interact with other people. Each syllogism was inserted in a dialogue that marked the end of a short scenario, the major premise being uttered by one character and the conclusion by the other character. Eight different scenarios were created and all the syllogisms appeared in each of them. They had the same structure as the following one:

> On his way to school in the morning, Pierre always walks across the park where one can observe shrubs, trees, and many kinds of flowers: roses, asters, and tulips. On arriving at school, he told his schoolmate Marie:
> "In the park, all the flowers are frozen".

And Marie replied:

"Therefore in the park, all the roses are frozen".

Questions:

1) What flowers are there in the park?........................

2) What did Pierre tell Marie?........................

3) What did Marie reply to Pierre?......................

4) What do you think of Marie's reply? (Choose the correct answer and complete the sentence)

a) What Marie said is certainly true because.....................

b) What Marie said is certainly false because.....................

c) What Marie said may be true or false, one cannot know because..............

The first comprehension question aimed to make sure that the child identified the set and its subsets (which provide the implicit premise, *all roses are flowers)* even though these were presented in the scenario. Because in Sloman's (1998) data the likelihood for the relation to be heeded increases with the similarity between categories (and there is also some evidence of this in Cherniak, 1984) we used set relations that provide high similarity or typicality such as rose/flower, screwdriver/tool, or poodle/dog. The second and third questions aimed to make sure that the child understood the premise (uttered by the first character, here Pierre) and the conclusion (uttered by the second character, Marie). The fourth question aimed to get the child's judgment about the logical status of the conclusion.

The particular quantifier (English "some") was rendered in French by "il y a" instead of "certains" which is more commonly used in reasoning tasks. This is because the scalar implicature ("not all") is less likely to be triggered by the former than by the latter.

In studies of syllogistic reasoning in which a conclusion is presented for evaluation, the conclusion is either true or indeterminate, so that there are only two options, such as follows/does not follow, or true/indeterminate. Because in the set of 16 syllogisms under study, 11 are valid and 5 invalid, performance could have been enhanced artificially if participants had a bias towards answering *true*. To eliminate this possibility we used false conclusions while adding a *false* option to all the trials. For each valid syllogism the negation of the conclusion was presented on half of the trials (the other half having the usual affirmative conclusion). The use of a third option had the additional advantage that it provided a second test of the performance for the valid syllogisms. The example given above is one of a valid syllogism (AA-1: all M P, all S M) with an affirmative conclusion. Its negative counterpart had the following conclusion: "Therefore in the park some roses are not frozen".

Because invalid syllogisms have no determinate conclusion the question of what sentence to present for evaluation arose. We presented the quantified sentence which the literature indicates to be the most frequently produced or chosen by participants [note 5].

Given that the critical set of 16 syllogisms consists of all the syllogisms with an A minor premise (11 valid and 5 invalid) these constraints define 11 valid syllogisms in two versions (affirmative and negative conclusion) and 5 invalid syllogisms in one version, hence a total of 27 different arguments that will be called *problems*. Consequently the number of stimuli was 27 (problems) x 8 (scenarios) = 216.

3.1.3. Administration

For this and the next experiment the task was administered at school during a French language class in the course of a routine exercise of text comprehension presented by the teacher. The children solved only one problem per session. There were eight sessions spread over four to eight weeks, so that each child solved eight problems, one in each of the scenarios. The 16 problems were divided into two blocks of eight, each one administered to one half of the children. The figures and moods were roughly equally represented within and between blocks. This resulted in a pseudo-random order of presentation within each block. Each problem was presented on a sheet of paper with a space for the children to write their answers, which they did at their own pace [note 6].

Before the first session proper there was a preliminary session that aimed to explain how to choose the options of the last question. Four questions were asked and the correct answer explained. Knowing that Mr. Dupont has a dog, the teacher asked whether (a) it is right to say that Mr. Dupont has a four-legged animal: Yes, this is certainly true because a dog has four legs; (b) it is right to say that Mr. Dupont has an animal that meows? No, this is certainly false because a dog does not meow. (c) Knowing that Mr. Dupont has a four-legged pet, is it right to say that Mr. Dupont has a dog? No, one cannot be sure because there are other four-legged pets. (d) Knowing that Antoine has marbles in all colours, is it right to say that he has red marbles? Yes, even though he has marbles in other colours, he does have red marbles.

3.1.4. Evaluation of the participants' answers

For each problem, the child chose one option *(true; false; may be true or false)* and wrote up a justification for this choice. The request for a justification aimed to increase the accuracy of the measurement. In principle, performance could be assessed based only on the option chosen but the result of an exploratory study suggested that this could be misleading: there were cases in which the choice of a correct option was made for wrong reasons, as indicated by incorrect justifications. Conversely there were cases where the choice of an incorrect option was made for good reasons; this occurred when the child had her own interpretation of the problem for pragmatic reasons that will be detailed below, so that a logically correct option could consistently be expressed by one of the other two options (e.g., choosing *false* instead of *may be true or false).* Consequently, calling the combination of the option and its justification a *response*, in this and the next experiment a decision about the correctness or incorrectness of a response was made only after an examination of the internal consistency between an option and its justification. We indicate below the various cases that occurred and give detailed examples of how the responses were assessed in Appendix B.

The correct responses. There are two cases depending on the logical correctness of the option chosen.

First case. The option is logically correct and the justification supports it or is just compatible with it. This constituted 90% of the correct responses.

Second case. The option is logically incorrect but pragmatically licensed under an interpretation of the question or of the sentences (premise or conclusion) that is clearly supported by the justification (see examples in Appendix B). The child is actually consistent. Such interpretations led children to choose the option *false* instead of *may be true or false* on 22 observations (6% of the correct responses); or the option *may be true or false* instead of *true* on 12 observations (3% of the correct responses).

Finally, there was a very small number (N = 5) of obvious cases of a slip of the pen.

The incorrect responses. Again there are two cases depending on the logical correctness of the option chosen.

First case. The option is logically incorrect and the justification does not constitute an argument that clearly supports the correct option. This constituted 85% of the incorrect responses.

Second case. The option is logically correct but either the justification is not based on logical grounds or it clearly supports another option. There were 23 observations of this type (15% of the incorrect responses).

Finally, a number of observations were eliminated because it could be inferred from the analysis of the justifications that the syllogism actually solved did not coincide with the one expected to be solved. In a few cases (N = 6) a particular premise was mistaken for a universal premise; the reason is that in some contexts the formulation chosen for the particular quantifier (French *il y a)* may be construed as an exhaustive description of what there is, instead of quantifying the subject. In the other cases (N = 10) the justification indicated that the participant denied the implicit inclusion relation of the subset in the superset. Most cases happened with the cake shop scenario for which the child stated that a tart is not a pastry, which changed the minor premise from A (all) to E (no). After discarding these 16 observations there remained 540 observations on which the results are based, yielding an average of 20 observations per syllogism.

3.2. Results and Discussion

The rate of correct responses was 71.3%, ranging from a minimum of 41% (for OA-4: some P not M, all M S) to a maximum of 92% (for AA-3: all M P, all M S). Table 1 indicates

Table 1. *Experiment 1. Percentage of correct responses for each syllogistic form.*

syllogism				syllogism		
AA1: all M P, all S M	+ 89	(75)		IA1°: some M P, all S M	57	(15)
	- 90					
AA2°: all P M, all SM	70	(40)		IA2°: some P M, all S M	47	(30)
AA3: all MP, all M S	+ 95	(20)		IA3: some M P, all M S	+ 85	(75)
	- 89				- 79	
AA4: all P M, all M S	+ 75	(85)		IA4: some P M, all M S	+ 65	(90)
	- 80				- 84	
EA1: no M P, all S M	+ 77	(80)		OA1°: some M not P, all SM	52	(20)
	- 68					
EA2: no P M, all S M	+ 71	(80)		OA2: some P not M, all S M	+ 50	(40)
	- 81				- 52	
EA3: no M P, all M S	+ 68	(15)		OA3: some M not P, all M S	+ 68	(40)
	- 81				- 86	
EA4: no P M, all M S	+ 55	(5)		OA4°: some P not M, all MS	41	(20)
	- 65					

Note. For the valid syllogisms + and - indicate whether the conclusion to evaluate was true or false. Median values resulting from a meta-analysis of studies with adults are in parentheses (see text). °indicates the invalid forms.

the distribution of correct responses in percent rounded to the nearest unit. The differences between the positive and the negative conclusions lie between 1% (for AA-1: all M P, all S M) and 19% (for IA-4: some P M, all M S) with an average of 9.5%, indicating a good reliability of the measure.

Before turning to the main experiment, we make a first appraisal of children's performance on this task and then consider some possible improvement on the method.

To appraise performance we make a comparison with previous results on the traditional task. Unluckily, data for children are very scarce. Johnson-Laird, Oakhill & Bull (1986) studied children in an age range (9 to 11 years old) that overlaps our participants' but they used 64 pairs of premises (of which ours are only a subset) and did not report results for each syllogism separately. The mean correct was about 29%, to be compared with adults' performance with the same material reported to be equal to 59%. Detailed data are provided in another study (Bara, Bucciarelli & Johnson-Laird, 1995). Children aged 9 to 10 years and adults solved a subset of 28 syllogisms. To allow a comparison we use the 10 syllogisms common to these 28 syllogisms and to the 27 problems of the present study. The rate of success was 46.5% for children (57.5% for adults) in their study whereas it reached 78% for our children solving natural syllogisms. This is a much superior performance but the response formats differ: the conclusion was produced by the children in the two studies mentioned, whereas it was evaluated in the present experiment, which may have favoured our participants.

Another way to appraise performance is to perform meta-analyses and consider the order of magnitude of the success rates observed for a century of research on syllogistic reasoning. Restricting ourselves to the 16 syllogistic forms that we have used, adults' average performance on these 16 forms varies (across ten studies in English or in French known to the present authors) from a lower value of 32% to an upper value of 67% with a median of 45%. (These values are close to the average values provided by Khemlani and Johnson-Laird's

(2012) meta-analysis based on six studies that overlap only two of the studies used in our data). The rather wide variability across studies is due to a number of factors that have been shown to affect performance such as the materials (the terms may refer to concrete or abstract categories) and the response format (the response may be produced or result from a multiple choice). Given that the present rate of success for the 16 forms was 66.6% (considering only the problems with a true conclusion, but this figure rises to 71.3% when the problems with a false conclusion are added), we get a telling assessment of our participants' performance: 11 years old children solving natural syllogisms performed exactly as the adult's *best* performance in solving traditional syllogisms; in other words they outperformed the adults. This global appraisal still holds when one considers the syllogisms separately. For each of them the median of the success rates provided by the meta-analysis (indicated in Table 1) was taken as a norm of comparison. Chi-square tests performed at a level of $p < .05$ showed that, with one exception, performance was either significantly higher than the norm (for ten problems) or did not differ significantly (for five problems).

Another result is noteworthy and can best be appreciated if one considers that the adults' rates of success just mentioned are averaged across logical form (figure and mood) which is by far the most important factor of variability: for the easiest forms the rate of success is above two thirds and can reach 100%, whereas for the hardest forms the rate is less than one quarter and can collapse to 0%. Keeping this in mind while examining Table 1, it appears that on the most notoriously difficult syllogisms (AA-3: all M P, all M S; EA-3: no M P, all M S; and especially EA-4: no P M, all M S) our participants' performance is incommensurate with the rate commonly observed when adults solve traditional syllogisms. A related point is that the syllogisms known to be the easiest lend themselves to a ceiling effect: performance was about the same on AA-1 (all M P, all S M), AA-4 (all P M, all M S), EA-1 (no M P, all S M), EA-2 (no P M, all S M), and especially IA-4 (some P M, all M S).

Bringing children's performance to or above the level of adults solving traditional syllogisms is a remarkable achievement. But one may wonder whether our method of evaluation of children's responses was not too lax. That is, in taking into account both the option and the justification to define a correct response, we were led to consider as correct a number of answers that would have been considered as erroneous if we had not done so. If these answers are reclassified as erroneous, the overall performance remains very high: 65% instead of 71.3%. But note that by parity the symmetrical cases (correct options leading to erroneous responses after taking into account the justifications) should be reclassified as correct. When this is done the percent correct drops only to 69%. In brief, the high performance observed cannot be attributed to our method of evaluation.

As usual the validity status introduces disparity: children were 75.1% correct on valid syllogisms but only 54% on invalid syllogisms. Nevertheless this performance is also remarkable because we proposed as a conclusion a logically incorrect proposition that is highly attractive (it is reminded that we chose the most frequent answer) to which children were able to resist more than half the time. Had we proposed a less frequent answer, in all likelihood this would have been less often endorsed, leading to more frequent correct answers.

In sum, we have obtained data showing that 11-year-old children are proficient in solving natural syllogisms and attain a high performance on this task —by the standards of the robust results reported in the literature on syllogistic reasoning with syllogisms whose premises express arbitrary or contingent inclusion relations.

Even though these results strongly support our expectations, we must be cautious and examine whether in addition to the main manipulation, which consisted of replacing contingent categories with natural categories within a dialogue, some features of our paradigm could have acted as confounding factors. The observed high performance could be attributed to one or several of the following features: the conditions of administration that put low demand on the children as they had to solve only one syllogism per session; the request for a justification that could have induced the children to reflect more on the options; the existence of a scenario that helped motivate the children; and the preparatory questions. But note that the first three features of our procedure could only help suppress factors potentially detrimental and could not have enhanced performance artificially. The last feature is to be considered because the instances chosen to explain when a conclusion follows logically (it is certainly true that Mr. Dupont has a four-legged animal knowing that he owns a dog) and when it is indeterminate (one cannot be sure that Mr. Dupont owns a dog knowing that he has a four-legged animal) were based on inclusion relations. These preliminary questions and their answers could have constituted some kind of logical training transferable to the syllogistic task proper. Even if it seems doubtful that four questions are sufficient to have such an effect, in the main experiment the questions used in the preliminary session were based on transitivity instead of inclusion.

4. Experiment 2

4.1. Method

4.1.1. Material and design

The methodology developed for the first experiment was applied to the main study which consisted of a comparison between two conditions that differed by the minor premise of the syllogism: The *contingent* condition was a control condition that presented a standard artificial syllogism with two explicit premises, one of which (the minor premise) stated a contingent inclusion relation (e. g., *all the potted plants are flowers,* as opposed to an implicit natural inclusion relation such as *all roses are flowers).*

The material was basically the same as in the first experiment, with a few modifications. The most important change was introduced to take into account the main superficial difference between the arguments in the two experimental conditions, namely the absence of an explicit minor premise in the natural condition and its presence in the standard (control) condition. Because in the control condition two pieces of information (the two premises) were given, participants in this condition could have been disadvantaged as they would have had to read a text that is longer than the text of the natural syllogism. To avoid this, a sentence of about the same length as the first premise of the contingent syllogism was added to its natural counterpart; it was relevant in the scenario but logically irrelevant to draw a conclusion. A few other changes were made as follows.

To avoid the interpretation of the particular quantifier as a universal which occurred occasionally, the more usual quantifier (French "certains") was used. Two, to limit the number of items, the problems with a false conclusion were suppressed and the number of options was accordingly reduced from three to two. Note that in the first experiment the rate of correct responses for the true conclusions was lower by 5% as compared to the false conclusions, so that presenting only the true conclusion problems could, if anything, result in a lower performance. Three, while the content of the scenarios was maintained, a few changes of hyponym were made to make all the subset to set inclusions indisputable. Four, there were slight modifications in the questions. The three initial comprehension questions were reduced to two because the first one proved to be useless in the first experiment as it was answered correctly 100% of the time. The children had to answer only two comprehension questions (what did [the first character] tell [the second character]? and what did [the second character] reply?) by underlining the relevant sentences.

The same scenarios were used in the two conditions. Two sentences were uttered by the first character. For instance, for the AA-1 syllogism in the park scenario, after "on arriving at school, he told his schoolmate Marie", the scenario read:

"In the park, all the flowers are frozen",

and then he said:

"all the potted plants are flowers".

And Marie replied:

"Therefore in the park, all the potted plants are frozen".

In the *natural* condition the dialogue was as follows:

"In the park, all the trees are losing their leaves",

and then he said:

"In the park, all the flowers are frozen."

And Marie replied:

"Therefore in the park, all the roses are frozen".

Because there was irrelevant information in the natural condition, the logic question in both conditions was as follows:

Using what seems important to you in [the first character's] utterance, give your opinion regarding [the second character's] reply. Think it over carefully, choose an answer (a or b) and explain your choice.

a) What [the second character] has replied is certainly true because...

b) What [the second character] has replied may be true or false, one cannot know because..

4.1.2. Participants and administration

The participants were 96 children (44 girls) attending a primary school in five small French cities. Their mean age was 10 years and six months (sd = 0.4 months). The procedure for the administration was the same as in the first experiment with the exception that each child was allocated to one of the two experimental conditions.

For the preliminary session, transitive inferences were used. For instance, given that Zoe is older than Nathan and Nathan is older than Laura, the children were asked whether Zoe is older than Laura to illustrate the answer *certainly true.* And, given that Zoe is older than Laura and Nathan is older than Laura, they were asked whether Zoe is older than Nathan to illustrate the answer *one cannot know.* A similar pair of questions were asked referring to the height of trees.

4.2. Results

Like in the first experiment, the method of analysis was based on the examination of the consistency between the option chosen and its justification. Table 2 indicates the distribution of correct responses in percent for the two conditions. The mean rate of correct responses was 72.5% for the natural condition (range: 45.8% - 95.8%), and 48.2% for the contingent condition (range: 12.5% - 79.2%), showing an absolute superiority of more than 24 percentage points in favour of the former. For all the syllogisms but one the success rate was higher in the natural condition than in the contingent condition, showing a main effect significant at the level of $p < 10^{-3}$ (sign test). Considering the syllogisms individually, nine

Table 2. *Experiment 2. Percentage of correct responses for each syllogistic form in the two conditions.*

syllogism	condition		syllogism	condition	
	natural	contingent		natural	contingent
AA1: all M P, all S M	83.3	70.8	IA1°: some M P, all S M	50.0*	20.8
AA2°: all P M, all S M	75.0*	12.5	IA2°: some P M, all S M	54.2*	16.7
AA3: all M P, all M S	87.5	79.2	IA3: some M P, all M S	81.0*	57.1
AA4: all P M, all M S	83.3*	58.3	IA4: some P M, all M S	95.2	76.2
EA1: no M P, all S M	95.8*	62.5	OA1°: some M not P, all S M	58.4*	29.2
EA2: no P M, all S M	56.5	50.0	OA2: some P not M, all S M	62.5*	20.8
EA3: no M P, all M S	83.3*	41.7	OA3: some M not P, all M S	90.5	76.2
EA4: no P M, all M S	58.3	75.0	OA4°: some P not M, all MS	45.8	25.0

° invalid syllogisms * $p < .05$

differences were significant at the level of $p < .05$ (chi-squares or Fisher tests). Out of the remaining six non-significant differences, four coincided with a ceiling effect (percentages > 70% in the contingent condition). In brief, the performance on the natural syllogisms was reliably superior to the performance on the contingent syllogisms by about 24 percentage points (corresponding to a rate of correct responses superior by one half).

Considering both conditions, like in the first experiment 90% of the correct responses were right away consistent responses and the remaining 10% were responses in which the justification showed logical comprehension while the apparently incorrect option was pragmatically justified. For the incorrect responses, 74% were right away inconsistent, while the remaining 26% were correct options accompanied with illogical justifications. We checked whether our method of evaluation could have spuriously increased the performance in any condition. This was not the case as the difference in performance remained identical after discarding the observations in which the final evaluation of the response mismatched the initial evaluation of the option as correct or incorrect.

5. General discussion

We found that our preadolescent participants were proficient in solving natural syllogisms as the rate of correct responses was about 70% in both experiments. The robustness of this result is remarkable as performance remained the same in spite of the differences that concerned the presence or the absence of an irrelevant premise, the polarity of the conclusion (affirmative or negative), the response format (two or three options), and the wording of the particular quantifier. In sum, the claim that there exist syllogistic arguments easily understood in daily life, provided the minor premise is an implicit inclusion relation between categories stored in semantic memory, already supported by previous research with adults, is now reinforced by the present results which extend it to 10- to 11-year-old children [note 7]. This in turn bears out the ecthesis theory, that is, the use of multiple predication crucial to the solution, which predicts and explains the ease of comprehension of natural syllogisms. We now examine whether our theory explains all our data, and whether there are alternative explanations of the results.

5.1. Differences in Error Rate between Syllogisms and Prediction of Errors

The overall high rate of correct responses is an average value that conceals some sharp differences between problems, which an adequate theory should be able to account for.

The explanation of these differences is fairly straightforward. The 16 syllogisms can be classified on the basis of the subroutines they require. A first group of eight syllogisms use only <EXHAUST>: AA-1; AA-3; AA-4; EA-1; EA-3; IA-3; IA-4; OA-3. A second group of five syllogisms use <SUPPOSE>: AA-2; EA-2; EA-4; IA-2; OA-2; two syllogisms use <CO-REF>: IA-1, OA-1; and the last syllogism OA-4 uses both <CO-REF> and <SUPPOSE>. Applying <EXHAUST> requires reasoners to consider whether a characterization found for an individual member of a set can extend to all of them. We assume that the answer to this question, based on the direction of a subset to set relation, is fairly elementary. By comparison, the other two subroutines are more subtle. Applying <CO-REF> requires reasoners to be aware that two individuals with a common attribute may or may not be identical, and to subsequently refrain from attributing to a given member of a set a feature that may not be universally shared in the set. Applying <SUPPOSE> requires a suppositional step that may not be as readily available as <EXHAUST> (especially among preadolescents). Consequently we predict the first group to lead to better performance than the last two that need supplementary, and harder operations to execute. Finally the OA-4 problem that requires both <CO-REF> and <SUPPOSE> should be the most difficult.

The results are very clear-cut. Pooling across the two experiments, the rate of correct responses ranged from 78% to 90% (mean 83%) for the eight syllogisms of the < EXHAUST >

group. The other two groups yielded equivalent rates: from 51% to 72% (mean 61%) for the <SUPPOSE> group, 54% and 55% for the two <CO-REF> syllogisms respectively, while OA-4 had the lowest rate (44%) as expected. In sum, the general order of performance across groups was as expected, and the eight problems out of 16 predicted to yield the best performance were exactly as predicted (Fisher test, $p < 10^{-4}$).

Even though this result based on a quantitative analysis is remarkable, a complementary analysis can be considered. This analysis is qualitative and aims to make predictions of erroneous answers for each syllogism individually. However, the detection of erroneous conclusions requires a response format different from the format we used. Indeed, the erroneous answers must have a chance to be produced or chosen by the participants (e.g., by using a production task or a multiple choice), which our response format built in terms of *true* (or *false*) and *one cannot know* did not allow. For this reason, we will limit ourselves to just a few examples that illustrate how the theory can make such predictions. The method is simple: to predict the errors it suffices to examine which conclusions are produced by a wrong application of the subroutines. For example, the incorrect application of <EXHAUST> can result in the wrong quantifier (universal instead of particular, and vice-versa). Take AA-3 (all M P, all M S) for which the reasoner who has attributed P to every MS tests by <EXHAUST> whether this applies to every S. Failure to realize that this is not the case leads to a positive answer to the question and to the erroneous conclusion *all S are P.* Or to take another example, failure to apply <CO-REF> to IA-1 (some M P, all S M), that is, to realize that a selected SM need not coincide with a selected MP leads to their identification and to the erroneous conclusion *some S are P.* Similarly in applying <SUPPOSE>; consider IA-2 (some P M, all S M): the supposition that an SM individual is P which does not yield a contradiction, suggests a possible existence (but not necessary) and the erroneous conclusion *some P is S.* In brief, erroneous conclusions can be predicted qualitatively for each syllogism.

5.2. The characteristics of the task

Several characteristics of the task could have enhanced performance. They concern the format of presentation and the format of response, the use of an implicit premise, and the logical property of the minor premise.

5.2.1. The format of the task

The task made use of a scenario to present the syllogism, and the response format consisted of the evaluation of a single sentence. Both features are seldom used in the standard task and one might suspect that they acted as confounding factors responsible for the high performance. There are two main reasons to refute this hypothesis. One, the second experiment showed that performance was higher on natural syllogisms than on contingent syllogisms, keeping the use of a scenario and the response format constant. Two, the results of three experiments reported in Politzer (2011) show the superiority of natural syllogisms in tasks without scenarios and with different formats of response (3- and 5-option multiple choices).

Another issue related to the format is the risk for the conclusions to be subject to belief bias, hence for performance to be inflated by non-inferential processes. Note first that this concerns only children who would not pay heed to the strong specifying contextual information. But even for them a facilitating effect is unlikely to occur because there are about as many plausible conclusions (that is, true at large in the actual world, like *some roses are withered)* as implausible ones (false in the actual world, like *all the roses are withered),* so that the biases would cancel each other out. Finally, even if facilitation existed, it would apply to both conditions equally and leave the superiority of the natural condition unexplained.

5.2.2. The use of an implicit minor premise

The comparison between the contingent argument that has an explicit minor premise with the natural argument that does not have suggests more effort and attention in the former case due to the activity of reading and interpreting the minor premise which is absent in the latter case. We have taken this point into consideration in the second experiment in which the absence of an explicit minor premise was compensated with an additional (irrelevant) sentence. One might argue that a control should be made with a minor premise stating explicitly the natural inclusion relation. This was tested in Politzer (2011, experiment 2) and no difference with the implicit condition was observed.

5.2.3. The logical property of the minor premise

There is one characteristic of the natural task that may give rise to a more serious objection against the origin of the improvement in performance. Sentences universally quantified by *all* are indeterminate between strict and non-strict inclusion (Politzer et al., 2006). However, in natural syllogisms the implicit minor premise expresses a strict inclusion. Now suppose that some proportion of participants have a non-strict default interpretation (that is, an equivalence interpretation). Because the indeterminacy is removed, their equivalence interpretation is blocked. This may have the following consequence. Under the equivalence construal, the minor premise (whether it is all S are M or all M are S) identifies the M with the S, so that the reasoner is licensed to replace the M term in the first premise by S and adopt the resulting sentence relating S and P as the conclusion. Consider two instances of this. Given EA-3 (no M P, all M S), no M P yields no S P, which is an incorrect conclusion. Given EA-1 (no M P, all S M), no M P yields again no S P, which this time is the correct conclusion. For those participants whose equivalence interpretation is blocked these conclusions no longer follow by this easy replacement. This generalizes to all the syllogisms. A case by case analysis shows that for half of them (AA-2, AA-3, EA-3, EA-4, IA-1, IA-2, OA-1, OA-4) the production of an *incorrect* conclusion is suppressed, whereas for the other half (AA-1, AA-4, EA-1, EA-2, IA-3, IA-4, OA-2, OA-3) it is the production of a *correct* conclusion that is suppressed. Consequently, following the principle of the objection that assumes that a subgroup of participants are equivalence interpreters by default, and assuming that the mechanism just described is at least the predominant cause of the observed improvement in performance, an increase in the rate of correct responses is predicted for the first group of syllogisms, whereas for the second group it is predicted that no increase will occur (a decrease is likely). Although we did observe an increase for the first group, we observed an increase for the second group as well, not a decrease. In other words, the objection is definitely refuted, at least in the strong form just considered.

However, we can consider a weak form in which the equivalence interpretation could play a secondary role. It is suggested by the numerical data which indicate that the increase in performance was 26.6% for the first group of syllogisms but slightly less (22.0%) for the second one. We consider an additive model for the effects of the ecthesis and the equivalence interpretation. For the first group the increase due to ecthesis is augmented by a contribution due to equivalence (the sum adding up to 26.6%) while for the second group the increase due to ecthesis is diminished by a contribution due to equivalence (the sum adding up to 22.0%). This yields the average contribution of 24.3% for ecthesis and a contribution of 2.3% for equivalence, that is, an effect size of one tenth for the latter, which is negligible. We conclude that the hypothesis of an effect of the interpretation of the minor premise fails to explain the performance in the natural condition.

5.3. Relation to other theories

5.3.1. Argumentation theory

One remarkable result of this investigation is the good performance in the contingent condition of the second experiment (48.8% correct, a rate roughly equivalent to adults' median performance on the relevant problems in the traditional task mentioned earlier, and exactly equivalent to adults' performance in Bara et al., 1995). One reason for this suggests itself: it is to be found in the essential difference between the present task in the contingent condition and the standard task, which consists in the presentation of the argument in the form of a dialogue. From the viewpoint of argumentation theory (Mercier, 2011; Mercier & Sperber, 2011) the fact that one character in a dialogue uttered the conclusion to be assessed by the participants invited them to simulate the other character's role and exert their epistemic vigilance (Sperber & al., 2010), which there is no reason or incentive to do in the standard laboratory task. From this point of view, it is not the present response format that enhanced performance; rather, it is the standard task that fails to call up participants' logical skills. Of course, the natural group also benefited from the dialogic presentation (and possibly, like the contingent group, from the request for justifications which could have acted as a confounding factor). But this in no way affects the ecthesis theory, as shown by the second experiment in which the natural group definitely outperformed the contingent group. This confirms a similar superiority for the natural syllogisms observed in adults (Politzer, 2011). It would seem that the dialogic presentation raised children's performance to the adult's average level, and that the use of natural syllogisms raised this latter level to the adult's top level. This suggests an additive model, and we leave the test of its adequacy for further work.

5.3.2. Case identification

As mentioned above, the present study was designed to generalize the results of previous experiments (Politzer, 2011) administered to adults. This approach was inspired by the Aristotelian concept of ecthesis and its development by contemporary logicians (Thom, 1981) and developed independently of Stenning and collabotators' investigation of syllogistic reasoning based on case identification. The convergence is remarkable as it is apparent that the two approaches share essentially the same principle which attributes a pivotal role to the triple characterization of individuals at the competence and performance levels. They diverge only when considering people's basic competence. Stenning and collabotators relate performance on the syllogistic task to participants' interpretation of the task--those who adopt a sceptical attitude would tend to use classic logic and therefore show good performance, whereas the majority adopt a credulous attitude which leads them to using nonmonotonic logic. These participants have Gricean interpretation of sentences, some are abusive converters while some others are abusive non-converters, hence a typology of participants that can be related to performance, especially on invalid syllogisms. All of them search for individual characterization by looking for a "source premise" but the result differs as a function of premise interpretation (Stenning & Cox, 2006). The important point is the assumption that generally people can perform correctly on this task provided their interpretation allows them to adopt classic logic (and have the relevant computational competence). We have no doubt that taking into account interpretive factors is indispensable to explain performance on reasoning tasks. We have made extensive use of pragmatic theory to explain performance on various tasks (e.g., Politzer, 1986, 2004; Politzer & Macchi, 2000) including standard syllogisms (Politzer, 1990) but we do not think that the claim to the effect that a majority of people could solve standard syllogisms proficiently in the absence of interpretive difficulties has received direct substantial support. In sum, the objectives differ. While Stenning and collaborators aim to explain standard syllogistic reasoning, we aim to show what people are capable of when it comes to solve syllogisms, namely which syllogisms they can solve (the natural ones), and how. It is important to notice that the agreement between the two approaches (viz., regarding the identification of a critical individual as a theoretical concept that is crucial logically and psychologically) is much more fundamental than the disagreement. Our experiments were not designed to carry out a test between the two but this is certainly worth putting on the agenda.

5.3.3. Could theories of the standard syllogistic task explain performance?

At first sight, it is reasonable to assume, as a general principle, that already known information is processed more effectively than new information, so that any well-developed theory of standard syllogistic reasoning could predict better performance. However, after some reflection, this seems to be a superficial view and it is not obvious that it applies to the specific case in hand. Note first that one could make an opposite claim: the implicit premise must be retrieved from long-term memory, whereas the explicit premise is always available, being in a written form which provides an external prop. More essentially, going into the details of the mechanisms of the main theories of standard syllogistic reasoning, it is apparent that it is in the stage of abstract information processing, after the relevant features of the premises (quantifier, subject, predicate, figure) have been encoded, that working memory is heavily solicited, and this has little to do with the familiarity or entrenchment of one of the premises in semantic memory. Take for instance rule-based theories. An A premise is encoded as *"for all x, if x is M then x is P"* whether the premise is stated as *all roses are prizewinners* or *all roses are flowers*. Similarly, Chater and Oaksford's (1999) probabilistic model relies on heuristics that take into consideration quantifiers, their relative informativeness, and the coincidence between subject and end term but right away disregard the meaning of the terms which play no different role than that of abstract symbols. In brief, theories of syllogistic reasoning cannot attribute the results to familiarity with the content or storage in semantic memory. This conclusion applies also to another prominent theory, namely mental models, which assumes that models for quantified sentences are made of abstract tokens that represent members of sets: this is the case whether or not the quantified relations are stored in semantic memory. [note 8]

If world knowledge is not at the origin of the improvement in performance, maybe mental model theory could offer an explanation linked to the logical property of the minor premise, namely the strict inclusion relation (discussed in section 5.2.3). This idea leads one to examine what modification in the construction of the models of the premises is necessary to accommodate this specific feature, and whether this would result in a simplified representation allowing a greater ease of resolution. To do this, it seems natural to exploit the theory's notion of semantic modulation (Johnson-Laird and Byrne, 2002) in the constitution of models. In the present case, world knowledge suggests that in addition to *all S are M,* it is also the case that *some M are not S* (and similarly for their respective converses). Introducing the representation of this additional premise would alter drastically the models and increase considerably the difficulty of their constitution, making the task much more difficult instead of facilitating it. Therefore this possibility must be eliminated. There remains another possibility that focuses on the representation of A sentences. To represent *all S are M,* the theory puts the S tokens, but not the M tokens, between square brackets to indicate that the S, but not the M, are represented exhaustively (Bucciarelli and Johnson-Laird, 1999). So the representation could be modified by adding one M token. This is a limited alteration (but one that, if anything, renders the representation more complicated) that leaves unchanged the subsequent operations on the models and consequently does not result in the kind of deep facilitation in processing that is needed to explain the improvement in performance. In other words, it seems that mental model theory cannot explain this improvement by the logical characteristic of the minor premise.

5.4. Scope of the study and perspective

Our general claim is that ecthesis, which is, at the computational level, an inferential process by which traditional syllogisms can be solved, is also at the algorithmic level an inferential process by which reasoners can in principle solve syllogisms on the condition that it should be brought about by some special feature, be it grammatical or semantic. A specific prediction (already successfully tested with adults) has been confirmed here with adolescents: it concerns those syllogisms, which we have called "natural syllogisms", that have the property of containing a universal affirmative premise stating an inclusion relation kept in semantic memory. The subset of syllogisms so defined contains most valid syllogisms, hence its practical importance. Another prediction mentioned above concerns singular syllogisms, in which the use of demonstratives proves efficient, and this applies to the remaining valid moods too (those with an E minor premise, see Politzer & Mercier, 2008). In sum, syllogisms are within the grasp of lay reasoners provided the ecthetic strategy can be triggered. Their investigation is worthwhile but this has so far been misdirected.

Indeed our results have far-reaching implications regarding the traditional laboratory task. We have mentioned earlier how mediocre the performance is. This obtains across materials, instructions and response formats. Besides the regular observation of a few biases and "effects" that are task specific, little knowledge has been gained from this task. Interestingly, several theoretical approaches resort partly or totally to non-logical heuristics to account for the data. All this points to the same conclusion: the task generally exceeds reasoners' logical capabilities. It teaches us hardly more than this: people are bad at solving this kind of formal arguments. Its use does not concern research on reasoning but rather research centered on a kind of puzzle, and how participants manage to get out of it. In fact it is not even clear to which target population the task could validly be administered. This is hardly surprising if one considers that this paradigm was designed more than a century ago (Störring, 1908) at a time when logicism was the dominant view —which lasted for the following six or seven decades— and categorical syllogisms were viewed as both a normative and a descriptive model of human reasoning.

In the last decade, psychological research on human reasoning has extended its interest toward informal reasoning and argumentation in natural settings. From this point of view, it is worth examining the differences between the arguments used in the traditional and the natural syllogistic tasks. This concerns their conditions of use, aim, surface structure, and source of information. In the classical task the argument comes in isolation, is aimless, has a conventional rigid surface structure, and its minor premise conveys new arbitrary information. In contrast, the natural argument typically is inserted in a dialogue, goal-oriented, its premises may not be explicitly stated and its conclusion may not come after the premises; and especially the minor premise consists of an inclusion relation that is already known. In sum, as this sharp opposition shows, the classical task is an artificial exercise that has no ecological validity. We conclude that, insofar as psychological research aims to explain reasoning occurring in daily life, it might be advisable to consider giving up a paradigm that is outdated, invalid, and what is worst, does not help reveal people's actual reasoning capabilities.

As a final remark, the use of what we call natural syllogisms in daily life is founded upon the inclusion relations that belong to our knowledge bases and on a need to argue. It is remarkable that this psychological result concurs with Aristotle's prominent interest in categorization and argumentation, which are the means and the end of his syllogistic.

References

Bara, B. G., Bucciarelli, M., & Johnson-Laird, P. N. (1995). Development of syllogistic reasoning. *American Journal of Psychology, 108* (2), 157-193.

Braine, M. D. S. (1998). Steps toward a mental predicate logic. In M. D. S. Braine & D. P. O'Brien (Eds.), *Mental logic* (pp. 273-332). Hillsdale, N. J.: Lawrence Erlbaum.

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science, 23,* 247-303.

Byrnes, J. P., & Overton, W. F. (1986). Reasoning about certainty and uncertainty in concrete, causal, and propositional contexts. *Developmental Psychology, 22,* 793-799.

Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology, 38,* 191-258.

Cherniak, C. (1984). Prototypicality and deductive reasoning. *Journal of Verbal Learning and verbal Behavior, 23,* 625-642.

Corcoran, J. (1974). Aristotle's natural deduction system. In J. Corcoran (Ed.), *Ancient logic and its modern interpretations* (pp. 85-131). Dordrecht: Reidel.

Ford, M. (1995). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition, 54,* 1-71.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review, 109,* 646-678. DOI: 10.1037//0033-295X.109.4.646

Johnson-Laird, P. N., Oakhill, J., & Bull, D. (1986). Children's syllogistic reasoning. *Quarterly Journal of Experimental Psychology, 38A,* 35-58.

Khemlani, S., & Johnson-Laird, P. N. (2012). Theories of the syllogism: A meta-analysis. *Psychological Bulletin, 138* (3), 427-457. DOI: 10.1037/a0026841

Kneale, W., & Kneale, M. (1962). *The development of logic.* Oxford: Clarendon Press.

Lear, J. (1980) *Aristotle and logical theory.* Cambridge: Cambridge University Press.

Lukasiewicz, J. (1957). *Aristotle's syllogistic from the standpoint of modern formal logic.* Oxford: Clarendon Press.

Markovits, H., Schleifer, M., & Fortier, L. (1989). Development of elementary reasoning in young children. *Developmental Psychology, 25,* 787-793.

Mercier, H. (2011). Reasoning serves argumentation in children. *Cognitive Development, 26, 177-191.* DOI: 10.1016/j.cogdev.2010.12.001

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34* (2)*, 57-74.* DOI: 10.1017/S0140525X10000968

Mignucci, M. (1991). Expository proofs in Aristotle's syllogistic. In H. Blumenthal & H. Robinson (Eds.), *Aristotle and the later tradition* (pp. 9-28). Oxford Studies in Ancient philosophy. Oxford: Clarendon Press.

Morris, A. K. (2000). Development of logical reasoning: Children's ability to verbally explain the nature of the distinction between logical and nonlogical forms of argument. *Developmental Psychology, 36,* 741-758. DOI: 10.1037//0012-1649.36.6.74I

Moshman, D., & Franks, B. A. (1986). Development of the concept of inferential validity. *Child Development, 57,* 153-165.

Overton, W. F. (1990). Competence and procedures: Constraints on the development of logical reasoning. In W. F. Overton (Ed.), *Reasoning, necessity, and logic: Developmental perspectives* (pp. 1-32). Hillsdale, N. J.: Lawrence Erlbaum.

Patzig, G. (1968). *Aristotle's theory of the syllogism.* Dordrecht: Reidel.

Politzer, G. (1986). Laws of language use and formal logic. *Journal of Psycholinguistic Research, 15,* 47-92.

Politzer, G. (1990). Non-logical solving of categorical syllogisms. In J. P. Caverni, J. M. Fabre, & M. Gonzales (Eds.), *Cognitive biases* (pp. 193-204). Amsterdam: North-Holland.

Politzer, G. (2004). Some precursors of current theories of syllogistic reasoning. In K. Manktelow & M.-C. Chung (Eds.), *Psychology of reasoning. Theoretical and historical perspectives* (pp. 214-240). Hove: Psychology Press.

Politzer, G. (2004). Reasoning, judgement and pragmatics. In I. N. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 94-115). Houndmills: Palgrave.

Politzer, G. (2011). Solving natural syllogisms. In K. Manktelow, D. Over & S Elqayam (Eds.), *The science of reason* (pp. 19-35). Hove: Psychology Press.

Politzer, G., & Macchi, L. (2000). Reasoning and pragmatics. *Mind and Society, 1,* 73-93.

Politzer, G., & Mercier, H. (2008). Solving categorical syllogisms with singular premises. *Thinking and Reasoning, 14* (4), 434-453. DOI: 10.1080/13546780802407271

Politzer, G., Van der Henst, J.-B., Delle Luche, C., & Noveck, I. (2006). The interpretation of classically quantified sentences. *Cognitive Science, 30* (4), 691-723.

Rose, L. E. (1968). *Aristotle's syllogistic.* Springfield, Il: Charles C. Thomas.

Sloman, S. A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology, 35,* 1-33.

Smiley, T. J. (1973). What is a syllogism? *Journal of Philosophical Logic, 2,* 136-154.

Smith, R. (1982). What is Aristotelian ecthesis? *History and Philosophy of Logic, 3,* 113-127.

Sperber, D., Clément, F., Heintz, C., Macaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind and Language, 25* (4), 359-393.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition.* London: Blackwell.

Stenning, K., & Cox, R. (2006). Reconnecting interpretation to reasoning through individual differences. *Quarterly Journal of Experimental Psychology, 59* (8), 1454-1483. DOI: 10.1080/17470210500198759

Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science, 19,* 97-140.

Stenning, K., & Yule, P. (1997). Image and language in human reasoning: A syllogistic illustration. *Cognitive Psychology, 34,* 109-159.

Störring, G. (1908). Experimentelle untersuchungen über einfache Schlussprozesse. [Experimental investigation on simple inferential processes]. *Archiv für die Gesamte Psychologie, 11,* 1-127.

Thom, P. (1981). *The syllogism.* München: Philosophia Verlag.

Footnotes

1. Because they contain an *all* premise (used to express the inclusion relation between categories) natural syllogisms constitute a subset of only 28 arguments. But these are the most useful, in the sense that most of the excluded forms are invalid, and conversely most invalid forms do not belong to them. Also natural syllogisms include all the valid syllogisms with the exception of those that are in the IE mood since these have a negative universal minor premise which conveys an exclusion relation instead of an inclusion relation.
2. We have reviewed only 16 syllogisms out of the 28 that are relevant. We need not consider the others because 12 pairs are redundant when one of the premises is unstated. For example, IA-1 (some M P, all S M) and AI-4 (all P M, some M S) collapse into a single pair (to see this, exchange P and S and the order of the premises in AI-4 which is irrelevant when it is unstated).
3. As mentioned above (see note 1) by definition the forms in the IE mood do not belong to natural syllogisms, even though they are valid. It is clear now that, more generally, the forms with an E minor premise lack the fundamental quality of natural syllogisms, which is to prompt ecthesis. To see this, consider IA-3 and IE-3 which share the major premise *some roses are prizewinners* and compare their implicit minor premise *all roses are flowers* and *no roses are tulips,* respectively. Although from both a logical and a psychological point of view a rose is a kind of flower, only from the logical point of view can we say that a rose is a kind of non-tulip. This stems from the fact that substituting *flower* for *rose* (or vice versa) involves the addition or subtraction of a relatively small number of properties (viz., the specific differences of *rose* with respect to *flower*) whereas substituting *non-tulip* for *rose* (or vice versa) involves a huge number of properties (viz., the properties of all the categories in the universe except the tulips).
4. One could argue that Sloman (1998) has provided data that run counter to this affirmation,

that is, reasoners may not always take into account the inclusion relation between categories. However, we view these data as inconclusive because, except in one or possibly two cases, the inclusion relations were not between fundamental categories, that is, lexicalized, developmentally precocious and stored in long-term memory such as *rose - flower*, but were rather elaborate conjunctive concepts such as *bank manager - white collar worker* which do not satisfy these criteria. The author himself alludes to this problem as a limitation of his results.

5. This is always the sentence obtained by substituting the end term of the minor premise for the middle term in the major premise. Take for instance IA-2: some P M, all S M instantiated as *some broken objects are tools, [all screwdrivers are tools]).* The substitution of S for M in the major premise yields *some P S: some broken objects are screwrivers* as a conclusion to evaluate.

6. One could fear that across a relatively long period some children might have been influenced, e.g., by conferring with each other. This was unlikely to occur as on any session the problems differed across participants. Also, the teachers were instructed to not comment on the problems and the forms were collected after completion of the task. Finally, there were slight and unsystematic increases and decreases in performance across sessions, which is incompatible with the objection.

7. From a developmental point of view, the results vindicate *a posteriori* the choice of the age range and of the methodology of written protocol analysis for this study. Indeed, we found many justifications that unambiguously exhibited a grasp of logical validity in our participants.

8. In common with ecthesis theory, mental model theory represents individuals. Indeed, ecthesis involves the selection and manipulation of an individual, and mental models for quantified sentences are based on the representation and manipulation of a few individuals.

But this is only a superficial resemblance because mental models are iconic whereas no such assumption is made in the exposition procedure.

9. One might be tempted to argue that by parity of argument the same participants should be reluctant to accept universal statements such as *all roses are flowers,* on the ground that *roses* are just *roses*, not *flowers* that include other subsets. But this would miss the role of the principle of relevance that guides the interpretation of the statement. In one case, the hearer attributes a communicative intention to a speaker in a dialogue; in the other case the hearer contemplates a factual (biological) statement and recovers its literal meaning.

Appendix A

The four quantified sentences are traditionally designated by A, E, I, and O:

A: all X are Y = universal affirmative

E: no X is Y = universal negative

I: some X are Y = particular affirmative

O: some X are not Y = particular negative

The term that appears in both premises of the syllogism is called the *middle term* M, the other two are called the *end terms*, P and S, hence the four combinations called *figures* numbered here following the logical tradition:

	1	2	3	4
major premise	M P	P M	M P	P M
minor premise	S M	S M	M S	M S

In a denomination such as, e. g., EAO-4 the three letters indicate the *mood*, that is, the first premise (the major premise) is an E sentence, the second sentence (the minor premise) is an A sentence and the conclusion is an O sentence; the number indicates the figure.

In psychological studies one often considers only the premises. By extension it is usual to call "syllogism" what actually is just a pair of premises, such as EA-4 which refers to *no P is M; all M are S,* and to restrict the indication of the mood to the first two letters, here EA.

Appendix B

Here are examples of how the responses were assessed in both experiments. For the sake of simplicity, we first provide an example of each case using the same syllogism, namely OA-1 (some M not P, all S M), which is invalid. Then we provide a second example of each case using a valid syllogism.

The correct responses.

First case. The option is logically correct and the justification supports it or is just compatible with it.

Example 1 (OA-1): *In the cupboard, some containers have no crack; therefore some bowls have no crack.* Option chosen: *may be true or false.*

- An instance of a justification that supports the option: *It is not specified which containers have no crack.* This shows full understanding of why the syllogism is indeterminate.
- An instance of a justification compatible with the option: *Not all containers have a crack.* This provides only a paraphrase of the premise but overall the response is acceptable as it contains no inconsistency.

Example 2: (AA-1): *In the shelter, all the dogs are ear-marked; therefore all the poodles are ear-marked.* Option chosen: *true.*

- Two instances of justifications that support this option: (i) *poodles are dogs.* This makes explicit the implicit premise. The majority of the justifications in this category are of this type; (ii) *all the dogs are ear-marked and poodles are dogs.* This states the two premises, including the implicit one, showing more explicitly perfect understanding of the syllogism. By contrast, a justification just compatible with the option would state that *all the dogs are ear-marked* (a mere repetition of the major premise).

Second case. The option is logically incorrect but pragmatically licensed under an interpretation of the question or of the sentences (premise or conclusion) that is supported by the justification. The child is actually coherent.

Example 1 (OA-1): *In the basement, some tools are not broken; therefore some screwdrivers are not broken.* Option chosen: *false.* Consider the justification, *it is not only the screwdrivers that are not broken.* Could it be the case that, in answering *false*, the child means that all the screwdrivers are broken? Certainly not, as her justification is incompatible with this judgment. Rather, in a somewhat awkward manner, the child objects to focusing on the subset of screwdrivers, which results in overlooking the other subsets to which *tools* may refer. So it appears that the invalidity of the syllogism is very well understood. Yet the child chooses the option *false* instead of *may be true or false,* which is explainable by the pragmatics of the question. When a character in the scenario asserts a conclusion that can logically be true or false, children often aptly considered that the person was wrong. It seems to have been difficult for them, in the framework of an argument, to make the subtle distinction between characterizing an assertion as logically indeterminate (the conclusion), so opting for *may be true or false*, and estimating that the character who uttered this conclusion was wrong. They often adopted the second viewpoint, hence the choice of the *false* option. In the present case (and in equivalent cases) precedence is given to the justification over the choice of the option, leading to a reassessment of the response.

Example 2 (AA-4): *In the park, all the frozen vegetation are roses; therefore all the frozen vegetation are flowers.* Option chosen: *may be true or false.* Consider the justification: *that may be tulips, asters, etc.* The child mentions the complementary subsets that constitute the set of flowers, meaning in a concise manner that the conclusion may be false in case the word *flowers* refers to these subsets. There are other possible construals of this justification; however, the present construal reflects a widespread pragmatic phenomenon, pervasive in this and the other experiment, to the effect that children are reluctant to qualify a category by including it in a whole set (flowers) when the evidence consists of the inclusion in a subset (roses). This has a straightforward explanation in terms of relevance (Sperber & Wilson, 1995). What is required for the use of the hypernym to be appropriate (that is, optimally relevant) is that the quality be applicable to all the subsets. This construal is confirmed by a number of other observations. We keep the AA-4 syllogism in the same scenario to provide the next example in which the option chosen is *false* and the justification, *Peter says that in the park, it is only the roses that are frozen.* This clearly supports that one is not pragmatically licensed to generalize the quality *frozen* from the subset of roses to the set of flowers. There is a small group of justifications that are strikingly similar to this one across scenarios, *mutatis mutandis* [note 9].

The incorrect responses.

First case. The option is logically incorrect and the justification does not constitute an argument that clearly supports the correct option.

Example 1 (OA-1): *In the basement, some tools are not broken; therefore some screwdrivers are not broken.* Option chosen: *true.* Justification: *the screwdrivers belong to the tools that are not broken.* This statement does not support the conclusion because it is arbitrary (although it would support it if it were true).

Example 2 (IA-4): *In the cupboard, some of the cracked objects are bowls; therefore some of the cracked objects are containers.* Option chosen: *false*. Justification: *the cracked objects are not only bowls.* This is an arbitrary statement. Even if true it would be insufficient to bear on the truth or falsity of the conclusion.

Second case. The option is logically correct but either the justification is not based on logical grounds or it clearly supports another option.

Example 1 (OA-1): *In the park, some flowers are not frozen; therefore some roses are not frozen.* Option chosen: *may be true or false.* Justification: *It is true and false because in winter flowers are frozen and in summer flowers are normal.* The information added in the justification is factual. There is no logical explanation of the invalidity of the syllogism.

Example 2 (OA-2): *In the shelter, some ear-marked animals are not dogs; therefore some ear-marked animals are not poodles.* Option chosen: *true*. Justification: *some poodles are not ear-marked but some other dogs are not ear-marked.* These are additional assertions that do not follow from the premises and do not support the conclusion.

Acknowledgments

The authors are grateful to Phil Johnson-Laird, Keith Stenning, Sunny Khemlani, Hugo Mercier, and two anonymous reviewers for very helpful comments made on previous versions of this paper.

The authors thank the primary school teachers in Balagny sur Thérain, Cires les Mello, Mouy, Noailles, Saint-Leu d'Esserent, Ully Saint Georges for offering access to their classes.