ELSEVIER

Cognition 86 (2003) 223–251

COGNITION

www.elsevier.com/locate/cognit

Reasoning with quantifiers

Bart Geurts*

Department of Philosophy, University of Nijmegen, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands

Received 20 July 2001; received in revised form 14 May 2002; accepted 28 August 2002

Abstract

In the semantics of natural language, quantification may have received more attention than any other subject, and one of the main topics in psychological studies on deductive reasoning is syllogistic inference, which is just a restricted form of reasoning with quantifiers. But thus far the semantical and psychological enterprises have remained disconnected. This paper aims to show how our understanding of syllogistic reasoning may benefit from semantical research on quantification. I present a very simple logic that pivots on the monotonicity properties of quantified statements – properties that are known to be crucial not only to quantification but to a much wider range of semantical phenomena. This logic is shown to account for the experimental evidence available in the literature as well as for the data from a new experiment with cardinal quantifiers ("at least *n*" and "at most *n*"), which cannot be explained by any other theory of syllogistic reasoning. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Syllogistic reasoning; Semantics; Quantification; Generalized quantifiers

1. Introduction

In logic, inference and interpretation are always closely tied together. Consider, for example, the standard inference rules associated with conjunctive sentences:

$$\frac{\varphi \,\&\, \psi}{\varphi} \qquad \frac{\varphi \,\&\, \psi}{\psi} \qquad \textit{\&-exploitation} \qquad \frac{\begin{matrix}\varphi\\ \psi\end{matrix}}{\varphi \,\&\, \psi} \qquad \textit{\&-introduction}$$

* http://www.phil.kun.nl/tfl/bart.
E-mail address: bart.geurts@phil.kun.nl (B. Geurts).

PII: S0010-0277(02)00180-4

&-introduction allows a sentence of the form "φ & ψ" to be derived whenever φ and ψ are given, and &-exploitation licenses the derivation of either conjunct of "φ & ψ". Of course, this is what one should expect in view of the meaning of "&", which is that "φ & ψ" is false unless φ is true and ψ is true. In logic, the search for a system of inference is usually guided by a (possibly informal) construal of a set of logical constants, and inference rules are judged by the constraints they impose on the interpretation of such logical vocabulary as they involve. Not that such customs are particularly remarkable, for there clearly must be an intimate connection between the meaning of an expression and valid arguments which make essential use of that expression. What is remarkable is that such connections have not played an equally central part in the psychological study of deductive reasoning, and especially of syllogistic reasoning.

In the past two or three decades, the semantics of natural language has come into its own, and quantification may have received more attention than any other semantic topic. During the same period, the psychological study of deduction made great advances, too, and one of *its* central topics is syllogistic inference, which is just a restricted form of reasoning with quantifiers. Strangely enough, these two enterprises have remained disconnected so far. All current approaches to syllogistic reasoning are based on first-order mental representations, which encode quantified statements in terms of individuals. Such representations are unsuitable for dealing with many quantified statements (e.g. "Most A are B", "At least three A are B", etc.), but semanticists have developed a general framework which overcomes these problems, and it will be argued that this framework should be adopted in the psychology of reasoning, too.

The plan for this paper is as follows. I start out with a survey of the central facts concerning syllogistic reasoning, and then go on to discuss the main approaches to deductive inference, arguing that each is flawed in the same way: they all employ representational schemes that are inadequate in principle for dealing with natural-language quantification, and in this sense they are all *ad hoc*. I then turn to the interpretation of quantified expressions, and sketch the outlines of a general framework for dealing with quantification that is widely accepted in the field of natural-language semantics. Research within this framework has shown that certain logical properties are especially important to natural systems of quantification, and I contend that the very same properties go a long way to explain the peculiarities of syllogistic reasoning.

It bears emphasizing, perhaps, that the general view on syllogistic reasoning adopted here is not original with me. Indeed, the key ideas have a venerable ancestry and can be traced back in part to medieval times and partly to the founder of syllogistic logic, Aristotle. More recent developments in semantic theory have systematized these ideas and incorporated them in a much broader framework. Therefore, my objective is a modest one: to show that this view on quantification is relevant to the *psychology* of syllogistic inference, too.

2. Syllogistic reasoning

The syllogistic language is confined to four sentence types, or "moods":

All A are B	:	universal affirmative (A)
Some A are B	:	particular affirmative (I)
No A are B	:	universal negative (E)
Some A are not B	:	particular negative (O)

Although the scholastic labels A, I, E, O (from Latin "AffIrmo" and "nEgO") have all but ceased to be mnemonic, they are still widely used, and I will use them here, too. Most psychological studies on syllogistic reasoning have adopted the traditional definition according to which there are four classes of syllogisms, called "figures", which are determined by the arrangement of terms in the arguments' premisses; the order of the terms in the conclusion is always the same:[1]

Figure 1	*Figure 2*	*Figure 3*	*Figure 4*
B C	C B	B C	C B
A B	A B	B A	B A
A C	A C	A C	A C

Following standard practice, I will sometimes identify syllogisms by their moods and figures. Thus, "AE4O" stands for the syllogism of the fourth figure whose premisses are of type A and E, in that order, and whose conclusion is of type O; that is:

AE4O:
All C are B
No B are A
———
Some A are not C

There are 256 syllogistic arguments altogether, 24 of which are valid according to the canons of traditional syllogistic logic. Of these 24, only 15 are valid in modern predicate logic. The difference lies in the interpretation of the universal quantifiers "all" and "no". In predicate logic, sentences of the form "All A are B" or "No A are B" are vacuously true if the set of As is empty, and therefore the following inferences are not valid in predicate logic:

All A are B	No A are B
Some A are B	Some A are not B

Intuitively, these inferences appear to valid, however, and they are accepted as such by traditional logicians; therefore, for example, syllogism AE4O (displayed above) is valid in

[1] Sometimes, in the psychological literature, the premisses trade places, and conclusions of the form "C A" are admitted, as well; such variations can affect the outcome of an experiment, as we will see below.

traditional logic but not in predicate logic. It is the traditional notion of validity that is adopted in the psychological literature, and I will do so, too.[2] Hence, by default I will use the term "validity" to denote traditional validity.

Experimental investigations in syllogistic reasoning have explored a number of paradigms. In most cases, subjects were given two premisses and then asked either to choose from a list of possible conclusions (e.g. Dickstein, 1978, 1981) or simply to say what, if anything, followed from the premisses; the latter format has always been used by Johnson-Laird and his associates. Relatively few researchers (including Rips, 1994) have used evaluation tasks, asking subjects to decide whether a given argument was valid or not. By and large, these various paradigms yield the same results, but there are some differences, too, as we will presently see.

Chater and Oaksford (1999) compared five experimental studies that used the full set of 256 syllogisms: two by Dickstein (1978), two by Johnson-Laird and Steedman (1978), and one by Johnson-Laird and Bara (1984). Chater and Oaksford found that the results of these experiments are very similar, and that differences in design appear to have had little effect. In fact, the weakest correlation they observed was not between Dickstein's multiple-choice studies and Johnson-Laird's production studies, but between the two experiments by Johnson-Laird and Steedman (1978), which adopted the same paradigm. Chater and Oaksford computed the average number of times (weighted by sample size) each conclusion was drawn in the five studies just listed; their data are reproduced in Table 1.[3] None of the studies collated by Chater and Oaksford used an evaluation paradigm, but their figures are very much in agreement with those of Rips (1994), who did. There are just two salient exceptions: for the valid AAI syllogisms Rips obtained much higher scores than one should expect on the basis of Chater and Oaksford's meta-analysis, and the same holds for the valid AEO and EAO syllogisms. Almost certainly, these discrepancies are due to the fact that in an evaluation task alternative inferences never have to compete with each other: the paradigm allows subjects to endorse several conclusions for the same pair of premisses. By contrast, in the multiple-choice and production paradigms, a subject who judges AA4A to be valid (which it isn't) is thereby prevented from endorsing AA4I (which is valid). Furthermore, production and multiple-choice tasks may be more susceptible to interference from response factors (see below). So all things considered, the evaluation paradigm may be more suitable for gauging reasoning capabilities. In practice, however, this consideration is of minor importance, since all paradigms paint the same general picture.

Turning to the main trends in the data of Table 1, it is evident that logical validity is a

---

[2] However, this is not to say that I endorse the traditional notion of validity wholesale. Researchers in semantics and pragmatics generally agree that a universal quantifier *presupposes* that its domain is non-empty, and presupposition is not the same as logical consequence (see, for example, Geurts, 1999; Horn, 1989). Strictly speaking, therefore, the nine syllogisms that separate the two notions of validity have a different status, because they are contingent upon the presuppositions of "all" and "no". And this distinction is relevant from a psychological point of view, too, since the 15 syllogisms valid in predicate logic are easier than the ones that are valid in traditional logic only.

[3] Not shown in Table 1 is the percentage of times subjects concluded that "nothing follows" from a given pair of premisses. Such non-propositional conclusions (as they have been called) raise some highly problematic issues, but as far as I can tell none of these has any bearing on the principal tenets of this paper.

Table 1
Percentage of times each syllogistic conclusion was endorsed according to the meta-analysis by Chater and Oaksford (1999)[a]

premisses	conclusion				premisses	conclusion				premisses	conclusion			
& figure	A	I	E	O	*& figure*	A	I	E	O	*& figure*	A	I	E	O
AA1	**90**	**5**	0	0	AO1	1	6	1	57	IO1	3	4	1	30
AA2	58	8	1	1	AO2	0	6	3	**67**	IO2	1	5	4	37
AA3	57	**29**	0	0	AO3	0	10	0	66	IO3	0	9	1	29
AA4	75	**16**	1	1	AO4	0	5	3	72	IO4	0	5	1	44
AI1	0	**92**	3	3	OA1	0	3	3	68	OI1	4	6	0	35
AI2	0	57	3	11	OA2	0	11	5	56	OI2	0	8	3	35
AI3	1	**89**	1	3	OA3	0	15	3	**69**	OI3	1	9	1	31
AI4	0	71	0	1	OA4	1	3	6	27	OI4	3	8	2	29
IA1	0	72	0	6	II1	0	41	3	4	EE1	0	1	34	1
IA2	13	49	3	12	II2	1	42	3	3	EE2	3	3	14	3
IA3	2	**85**	1	4	II3	0	24	3	1	EE3	0	0	18	3
IA4	0	**91**	1	1	II4	0	42	0	1	EE4	0	3	31	1
AE1	0	3	59	6	IE1	1	1	22	16	EO1	1	8	8	23
AE2	0	0	**88**	**1**	IE2	0	0	39	30	EO2	0	13	7	11
AE3	0	1	61	13	IE3	0	1	30	33	EO3	0	0	9	28
AE4	0	3	**87**	**2**	IE4	0	1	28	44	EO4	0	5	8	12
EA1	0	1	**87**	**3**	EI1	0	5	15	**66**	OE1	1	0	14	5
EA2	0	0	**89**	**3**	EI2	1	1	21	**52**	OE2	0	8	11	16
EA3	0	0	64	**22**	EI3	0	6	15	**48**	OE3	0	5	12	18
EA4	1	3	61	**8**	EI4	0	2	32	**27**	OE4	0	19	9	14
										OO1	1	8	1	22
										OO2	0	16	5	10
										OO3	1	6	0	15
										OO4	1	4	1	25

A = all E = no
I = some O = some … not

[a] All figures have been rounded to the nearest integer; valid conclusions are shaded. Whenever two conclusions in the same row are valid, only the first one is valid in predicate logic.

major factor in determining performance on syllogistic tasks. To begin with, the most widely endorsed syllogisms tend to be valid. According to Chater and Oaksford's data, valid syllogisms are endorsed 51% of the time on average; invalid syllogisms, 11% of the time. Seventeen valid syllogisms score above the upper quartile point ($P_{0.75} = 16.5$), and the remaining seven are presumably undervalued because they are in competition with more popular syllogisms. In Rips' data the effect of validity is even clearer, for his first 22 syllogisms are all valid. Furthermore, high-frequency errors tend to occur in the vicinity of valid argument forms. For example, of the four AAA syllogisms, only one is valid (i.e. AA1A), and it is typically recognized as such; in fact, it is one of the easiest syllogisms. But the invalid AAA arguments appear to share in this popularity, and are endorsed well above average. Other clusters of arguments for which this holds are AII, IAI, AEE, EAE, AOO, and OAO. The upshot of these observations is that people are rather good at syllogistic reasoning: not only are valid arguments very often recognized as such, but when invalid arguments are considered to be valid, they are often identical, *modulo* figure, to valid arguments.

Of the four syllogistic sentence types, two license *conversion* whilst the other two do not: whereas it follows from "Some A are B" and "No A are B" that "Some B are A" and "No B are A", respectively, conversion is illegitimate in the case of "All A are B" and "Some A are not B". There are 256 pairs of syllogisms that are identical up to conversion, which is to say that they can be made to coincide by applying conversion to one of the premisses. For example, EA1E and EA2E are a conversion pair, because conversion applied to the first premiss of one yields the other:

	No B are C		No C are B
EA1E:	All A are B	EA2E:	All A are B
	No A are C		No A are C

From a logical point of view, one would expect people to perform the same on the members of a conversion pair, and this expectation is not disappointed: for the set of pairs $\langle \varphi, \psi \rangle$ which are intertranslatable by means of conversion alone, the correlation between the φs and the ψs is quite good ($r = 0.93$).[4] This observation lends additional credit to the notion that many of the errors in syllogistic reasoning are caused by *illicit* conversion. This idea, which goes back to Wilkins (1928) and has since been supported by many researchers, comes in a variety of incarnations, the most promising of which is that people have a certain tendency to infer "All B are A" from "All A are B", and likewise (though less importantly) "Some B are not A" from "Some A are not B". Illicit conversion accounts in large part for the errors made in syllogistic reasoning. In Chater and Oaksford's data, the 16 most frequently occurring erroneous inferences (endorsed 49–75% of the time) are all attributable to illicit conversion.

There is also independent evidence that people make conversion errors. In an experimental task with single-premiss arguments, about one-third of the participants will incorrectly convert "all" propositions, and for "some not" propositions about two-thirds will endorse conversion (Newstead & Griggs, 1983); performance on this task correlates with errors predicted by illicit conversion in syllogistic tasks (Newstead & Griggs, 1983). Further evidence for illicit conversion is the finding by Dickstein (1981) that more elaborate clarification of a syllogistic task substantially improves performance, but in a selective way: clarification is significantly less effective with conversion errors. Dickstein suggests that illicit conversion can be accounted for by a general preference for symmetric rela-

[4] As is customary in set theory, I count $\langle \varphi, \psi \rangle$ and $\langle \psi, \varphi \rangle$ as two pairs. If they are counted as one, the number of conversion pairs is halved, and we are faced with the question of what is supposed to correlate with what. This question is a delicate one, because the members of a conversion pair are connected by a symmetry relation, and there is no principled criterion for separating between the factors of the correlation. However, I expect that in practice this would not matter very much, because some random separations I tried out yielded scores that didn't deviate too much from the one quoted in the text.

tions, as demonstrated, for example, by Tsal (1977), and this explanation accords with the observation made above about the importance of *licit* conversion in syllogistic reasoning.

It has often been suggested that figure is a major factor affecting the difficulty of syllogistic arguments. Proponents of this view typically hold that syllogisms in figure 1 are the easiest, those in figure 4 are the hardest, while figures 2 and 3 lie in between (e.g. Evans, Newstead, & Byrne, 1993). Such claims are not entirely without foundation, but it is doubtful that they bear much weight. To begin with, we don't find straightforward empirical support for the proposition that easier syllogisms tend to be in the lower figures. Amongst the ten easiest syllogisms in Table 1, figures 2 and 4 are represented twice each, while figures 1 and 3 are represented three times each, which indicates already that the figural effect, such as it is, is not a particularly strong one. Furthermore, it is unlikely that a figural effect could be particularly forceful in view of the fact that syllogisms which are conversion pairs tend to evoke similar responses; as we have just seen any two syllogisms that are identical up to conversion tend to be equally difficult, though conversion always entails a change in figure.

Dickstein (1978) observed that quite a few of the early reports on figural effects were flawed for various methodological reasons, but chiefly because they confounded potential figural effects with the effects of other variables, such as validity and illicit conversion (cf. also Rips, 1994). Dickstein argued that when all these factors are taken into account, only 12 argument forms remain that are suitable for testing the effects of figure. From an experiment with this restricted sample, he inferred that "figure is a significant determinant of performance *within a specific subset of syllogisms*" (p. 80, emphasis added). The tacit implication is that, when all proper precautions have been taken, it cannot be established that figure is a major factor in syllogistic reasoning.

Rather more impressive results have been reported by Johnson-Laird and his associates, who showed that premisses of the form AB-BC encouraged subjects to produce AC conclusions, while premisses of the form BA-CB inclined them more towards CA conclusions; the other two ways of arranging terms in the premisses caused no clear preferences (Johnson-Laird & Bara, 1984; Johnson-Laird & Steedman, 1978). Johnson-Laird et al. attribute these results to two factors: the fact that they adopt a production paradigm, which forces subjects to formulate their own conclusions, and, relatedly, the fact that subjects are allowed to draw AC as well CA conclusions, which effectively doubles the standard set of syllogisms from 256 to 512.[5] However, a study by Wetherick and Gilhooly (1990), which had the same enlarged set of syllogisms but used a multiple-choice test instead of a production design, failed to replicate Johnson-Laird et al.'s findings, so it seems rather likely that it is the change of paradigm alone that is crucial.

It has often been suggested that figural effects are linguistic in origin (e.g. Rips, 1994; Wetherick & Gilhooly, 1990). The pattern found by Johnson-Laird et al. is that there is a preference for co-opting the subject of one of the premisses to fill the same grammatical slot in the conclusion. Given that the middle term (i.e. B) cannot figure in the conclusion, this simple rule predicts a preference for AC conclusions for figure 1 syllogisms, a preference for CA conclusions for figure 4 syllogisms, and no distinct preferences other-

[5] Note that allowing AC as well as CA conclusions yields the same collection of syllogisms as allowing the order of the premisses to vary.

wise: in figure 2, the A and C terms both act as subject in one of the premisses, and in figure 3 neither of them do. The rationale behind the subject rule is obviously pragmatic. An argument is just a special kind of discourse, and one of the main structural principles underlying natural discourse is topic continuity: you keep talking about Fred, say, until other topics become more urgent, and whatever is the topic of conversation will tend to act as grammatical subject; that is what subjects are for, pragmatically speaking. This line of thinking explains why figural effects are so much stronger in production experiments than in other designs: a subject who has to formulate his own conclusion perforce relies more on his linguistic competence than one who just is to say yes or no, or choose from a list of alternatives. However, this account also implies that figural effects tell us little about deductive reasoning *per se*.

To sum up the main findings of our brief empirical survey, we have seen that validity is one of the main factors shaping performance in syllogistic tasks, and that a good deal of the errors in syllogistic reasoning are due to illicit conversion of A ("all") and O ("some … not") propositions. Conversion is central in a more general way, too, since syllogisms that are conversion pairs strongly tend to elicit equivalent responses. Finally, I considered so-called "figural effects", arguing that they are less substantial than they have been claimed to be, and suggesting that such directionality effects as have been demonstrated are plausibly viewed as being linguistic not inferential in nature.

Although in its long history psychological research on syllogistic reasoning has accumulated a rich supply of experimental results, it must be noted that in a way its empirical base is rather narrow. Syllogistic logic covers only a fragment of predicate logic, and even predicate logic falls short of the plethora of deductive arguments expressible in natural language. Hence, there are ample opportunities for varying experimental materials, yet these opportunities have barely been explored. As far as I am aware, there have been no studies on cardinal quantifiers like "five", "at least six", "at most seven", etc., no studies on the role of negation in syllogistic reasoning, almost no studies on arguments with multiple quantifiers or relational predicates, and so on. Inevitably, this preoccupation with a handful of argument schemas demands its toll; for, as will be shown in the next section, *no* current theory can deal with certain extensions of the syllogistic language, some of which are downright trivial.

3. Psychological theories of syllogistic reasoning

Over the years, many theories about syllogistic reasoning have been proposed, the large majority of which fall into one of three families: logic-based approaches, mental-model theories, and heuristic theories. The theory to be presented below belongs to the first family. Existing accounts in the logic-based tradition are mostly based on natural deduction, which is a species of proof theory developed by Jaśkowski and Gentzen in the 1930s.[6] The inference rules for "&" cited in Section 1 are natural-deduction rules. The inference rules for the quantifiers are in the same format, as the following definitions for the

[6] Sundholm (1983) gives a nice introduction to natural deduction, and compares it to older Frege-Hilbert style systems.

universal quantifier illustrate; here "$\varphi[a/x]$" denotes the result of replacing all free occurrences of x in φ with the individual constant a:

$$\frac{\forall x\varphi}{\varphi[a/x]}\ \forall\textit{-exploitation} \qquad \frac{\varphi[a/x]}{\forall x\varphi}\ \forall\textit{-introduction}$$

The basic idea is straightforward: if $\forall x\varphi$ is given one may derive $\varphi[a/x]$, for any individual constant a. Conversely, if we have $\varphi[a/x]$, and a was chosen arbitrarily, then we may conclude $\forall x\varphi$.[7] Here is an example of a proof that involves both ∀-rules and both &-rules:

[1]	$\forall x$[Px & Qx]	*premiss*
[2]	Pa & Qa	*∀-exploitation applied to* [1]
[3]	Pa	*&-exploitation applied to* [2]
[4]	Qa	*&-exploitation applied to* [2]
[5]	Qa & Pa	*&-introduction applied to* [3] *and* [4]
[6]	$\forall x$[Qx & Px]	*∀-introduction applied to* [5]

In words: since "Px & Qx" holds for all individuals x, the same must be true for an arbitrary individual a, and as "Pa & Qa" is true, Pa and Qa are true, as well; therefore, "Qa & Pa" is true, and since a was chosen arbitrarily, this holds for all individuals, and therefore "$\forall x$[Qx & Px]" cannot fail to be true. This proof illustrates how, in order to arrive at a quantified conclusion, we have to reason about arbitrary *individuals*. This setup leads to implausible consequences. Intuitively, one is inclined to say that "$\forall x$[Qx & Px]" follows directly from "$\forall x$[Px & Qx]" by transposition of the conjuncts. But in standard systems of natural deduction, this cannot be proved directly, since the proof has to go via arbitrary individuals. The problem is quite general, and affects syllogistic arguments, too. Consider, for example, the archetypical syllogism known since medieval times as "Barbara":

$$\text{AA1A:}\quad \frac{\begin{array}{l}\text{All B are C}\\ \text{All A are B}\end{array}}{\text{All A are C}}$$

As this is one of the easiest syllogisms, one should expect the conclusion to follow more or less immediately, but a standard proof will take no less than seven steps. In an attempt to remedy this type of problem, psychologists working with natural deduction have introduced further rules of inference, which enable the reasoner to make inferential shortcuts. For example, Rips (1994) postulates, in effect, that Barbara *is* a rule of inference, which means, of course, that the argument above becomes provable in one step.

[7] To say that a was chosen arbitrarily means that a may neither occur in φ nor in the premisses (if any) from which φ was derived.

From a logical point of view, such shortcut rules are simply pointless, since they don't enhance the system's logical power; they merely help to shorten some of the proofs. From a psychological viewpoint, however, this strategy is more objectionable. On the one hand, it is blatantly stipulative. Rips' system, for example, predicts purely by fiat that AA1A and EA1E are easy, which of course diminishes the theory's explanatory potential. On the other hand, the introduction of shortcut rules raises the question of how deduction skills could ever develop. Consider the quantified version of "mental logic" by Braine (1998), which has no less than a dozen rules for reasoning with "all". The system is highly redundant, and its redundancy has to be of just the right sort: practically any other collection of shortcut rules would yield different predictions about the relative difficulty of deductive arguments. How might such a system be acquired? On what grounds does a child "decide" that AA1A is easier than most other syllogisms, and therefore merits a special shortcut rule? Or, if mental logic is innate, why did Mother Nature fit us with *this* particular set of rules? Such questions are never easy, to be sure, but theories such as Braine's and Rips' add a whole new dimension to what is already a hairy issue, simply because it is unclear why any cognitive system should be redundant in one particular way, as opposed to (literally) infinitely many others.

Logic-based approaches to deduction tend to be more powerful than others in that they generalize more easily beyond syllogistic argument forms. Since natural deduction was designed as a method of proof for full predicate logic, systems based on this method have the logical resources for dealing with propositions containing multiple quantifiers and many-place predicates. Yet, such systems have their limitations, too, especially if we want to use them as psychological models. To begin with, some quantifiers, such as "most" and "at least half of" aren't expressible in standard predicate logic at all. The reason for this, informally speaking, is that they refer to *sets*, not individuals. For example, "At least half of the foresters are vegetarians" states that the set of foresters who are vegetarians is not smaller than the set of foresters who aren't. Since predicate logic only allows for talk about individuals, it is not expressive enough for representing such sentences. By the same token, any system of inference that deals with quantifiers in terms of arbitrary individuals cannot handle arguments like the following, if Q is replaced with "most", say:

All vegetarians are teetotallers
Q foresters are vegetarians
―――――――――――――
Q foresters are teetotallers

Intuitively, it doesn't make much of a difference if Q stands for "all" or "most", and experimental evidence confirms this impression (Oaksford & Chater, 2001). However, even if it is granted that they are adequate for "all" and other universal quantifiers, current logic-based theories cannot be extended in a straightforward way to deal with "most" and its kin.

Even if a quantifier is expressible in predicate logic, the representations involved may be ill-suited for psychological purposes. Consider, for example, how "At least two foresters are teetotallers" goes over into predicate logic:

$\exists x \exists y[x \neq y \ \& \ \text{forester}(x) \ \& \ \text{teetotaller}(x) \ \& \ \text{forester}(y) \ \& \ \text{teetotaller}(y)]$

Since predicate logic doesn't offer the means for talking about sets, a rather cumbersome representation is called for: we have to introduce two individual variables and ensure that their values are distinct and that both stand for a forester as well as a teetotaller. The complexity of this representation is proportional to the rank of the cardinal that needs to be represented: "At least n A are B" requires n variables and $0 + \ldots + n - 1$ clauses of the form $x \neq y$. This peculiarity makes predicate logic an unlikely vehicle for reasoning with cardinal numbers. It entails, for example, that if we replace the Q in the argument above with "some" or "at least twenty", the former argument should be much easier than the latter. This is intuitively false, and the intuition is corroborated by experimental evidence (see Section 5 below).

I have argued that the mental representations used by logic-based theories of reasoning are unsatisfactory. They are incapable of capturing even the simplest non-standard quantifiers, the hurdle being that in predicate logic we cannot speak and reason about sets. This is what renders it flatly impossible to represent proportional quantifiers, such as "most" and "at least half of", and it is for the same reason that the predicate-logical way of dealing with cardinals yields representations that, though logically impeccable, are inadequate from a psychological point of view. And it is not only logic-based approaches that suffer from these problems: all extant theories of reasoning run into the same sort of trouble. To illustrate this, I will briefly discuss Johnson-Laird's mental-model framework and the probabilistic treatment of quantification proposed by Chater and Oaksford.

In the theory of mental models developed by Johnson-Laird et al. over the past two decades, quantified propositions are represented directly in terms of arbitrary individuals. For example, in the Bucciarelli and Johnson-Laird (1999) version of the theory, processing the premisses of AA1A (in non-canonical order) results in the suite of mental models shown in Table 2. Every line in a mental model represents an individual, so for the first premiss we have two individuals, which have the same properties, A and B. The second premiss gives rise to a similar model, which merges with the first so as to produce an integrated representation of the two premisses. This representation is a partial one; further information may be added, though not all possible extensions are allowed, with square

Table 2
Representation and integration according to the theory of mental models

	Mental model		
1st premiss: "All A are B"	[a]	b	
	[a]	b	
2nd premiss: "All B are C"		[b]	c
		[b]	c
Integrated model of the two premisses	[a]	[b]	c
	[a]	[b]	c
Extended model, i.e. counterexample against "All C are A"	[a]	[b]	c
	[a]	[b]	c
			c

brackets signalling that the property in question is "exhaustively represented". Once the argument's premisses have been encoded, preliminary conclusions can be formulated. In the case at hand, the integrated model verifies "All A are C" as well as "All C are A", but as these conclusions are based on a partial model they are not necessarily valid, and have to be tested. This is done by trying to refute each of the preliminary conclusions by a counterexample: an extended model in which the premisses are still true but the conclusion is false. Such a counterexample can be found for "All C are A" (as shown in the last row of Table 2) but not for "All A are C", so only the latter survives and is spelled out as the final conclusion.

One of the things critics of mental-model theory have complained about is that it is not quite clear what it *is*, not only because the theory has gone through so many revisions, but because its key tenets remain somewhat underspecified. Usually, a version of the mental-model theory comes with one or more computer implementations and a description of what these programs do, but in general this does not suffice to pin down exactly what mental models are. To illustrate, while the first model in Table 2 is said to represent the proposition "All A are B", we are also told that the model in the third row verifies the proposition "All C are A". The former claim suggests that individuals representing the subject term must be enclosed in square brackets, to encode that its representation is exhaustive; the latter suggests that this is not necessary. It is only because mental models lack an explicit semantics that such inconsistencies tend to go unnoticed.

Or consider the sentence "Two A are B". How can we represent this in a mental model? One might think that the first model of Table 2 is a plausible candidate, but this cannot be right, for two reasons at least. First, this model already represents the interpretation of "All A are B", which is patently not synonymous with "Two A are B". Secondly, if it takes two individuals to represent "two", then presumably it takes sixty individuals to represent "sixty", which gets us back to the same problem we discussed in connection with predicate-logical representations of cardinalities. This is not a coincidence, of course, since predicate logic and mental-model theory are both individual-based systems, which forswear reference to entities other than individuals. It is for this reason that the two accounts get into the same trouble with non-standard quantifiers.[8]

A rather different way of dealing with quantification is Chater and Oaksford's probabilistic semantics, which underlies their "probability heuristics model" of syllogistic reasoning (Chater & Oaksford, 1999; Oaksford & Chater, 2001). According to Chater and Oaksford, humans are geared towards reasoning with uncertainty; we were designed by evolution to reason not logically but probabilistically, hence it is quite reasonable to ask for a probabilistic interpretation of quantified expressions. And for some quantifiers at least such an interpretation is easy enough to provide. Thus, "All A are B" means, probabilistically speaking, that $P(\mathrm{B} \mid \mathrm{A}) = 1$, i.e. the conditional probability of B given A equals 1. Similarly, "No A are B" conveys that $P(\mathrm{B} \mid \mathrm{A}) = 0$, and "Some A are B" that $P(\mathrm{B} \mid \mathrm{A}) > 0$. As a matter of elementary probability theory, the conditional probabilities of the premisses of a syllogism will occasionally restrict the conditional probability of the

[8] Johnson-Laird, Byrne, and Tabossi (1989: 672) remark in passing that "[t]he model-based theory is readily extendible to deal with nonstandard quantification" (cf. also Johnson-Laird, 1983: 443). In view of the considerations adduced in the foregoing, however, such claims must be wrong.

conclusion, and whenever this happens, "logical" inferences can be drawn (shudder quotes are called for here, because the probabilistic account implies that there is nothing logical about such inferences). For example, if the conditional probability of the conclusion is 1, a proposition with "all" can be inferred.

One virtue of the probabilistic approach is that it affords a representation of proportional quantifiers, such as "most": according to Chater and Oaksford's definition, "Most A are B" means that $P(\mathrm{B} \mid \mathrm{A})$ is high though less than 1. In this respect, a probabilistic semantics is more expressive than the approaches we have considered before, but it is still not expressive enough. In general, propositions involving cardinal quantifiers cannot be translated into a probabilistic format. For example, if it is given that "Two A are B", we do not know what $P(\mathrm{B} \mid \mathrm{A})$ is unless it is also known how many As there are. It might be proposed, therefore, that "Two A are B" means that $P(\mathrm{B} \mid \mathrm{A}) = 2/\mathrm{card}(\mathrm{A})$ (where "card(A)" stands for the cardinality of the set of As). Thus, if there are five vegetarians altogether, "Two vegetarians are liberals" means that there is a 0.4 probability that a given vegetarian is a liberal. This proposal is up to a number of problems, the most obvious one being that it suffices for "Two vegetarians are liberals" to be true that there two liberal vegetarians; the grand total of vegetarians is irrelevant. In brief, going probabilistic is tantamount to claiming that *all* quantifiers are proportional, which is unintuitive for some (like "some") and demonstrably false for others (like the cardinals).

In the foregoing we have looked at each of the main approaches to deductive reasoning, and found that they all lack the expressive power for dealing with some quantifiers that would appear to be quite innocuous. I have focused my attention on cardinal expressions because they are common, simple, and yet manage to create problems of principle for all current theories. However, the trouble is not restricted to one or two types of quantifier; it is symptomatic of a much deeper problem, which is that all approaches to syllogistic reasoning are *ad hoc* from the vantage point of language understanding. It is a truism that solving a syllogistic task begins with an exercise in interpretation: how are the premisses (and, in some paradigms, the conclusion) to be construed? The range of possible answers to this question is restricted by what is known about the interpretation of quantified sentences, obviously, and as quantification happens to be one of the central topics in the field of natural-language semantics, one might expect semantic theorizing to have had at least some impact on psychological accounts of syllogistic reasoning. As it turns out, however, any such expectations will be disappointed: thus far the impact has been practically nil.

And it is not as if the semantic theory hadn't made any progress on the subject of quantification. On the contrary, it is widely agreed that the past two decades have taught us a great deal about this topic, and there is even a broad consensus on what is the best general framework for dealing with quantified expressions. In the following I will argue that this framework goes a long way to explain how people reason with quantifiers.

The plan for the remainder of this paper is as follows. Since my central claim is that a psychological account of syllogistic reasoning presupposes an adequate theory of interpretation, I start out by discussing the general framework for treating quantification that semanticists have settled on. Research within this framework has shown that there are certain logical properties that are especially relevant to natural-language quantifiers, and I present an inference system that capitalizes on these properties. The resulting model of syllogistic reasoning is motivated almost entirely by semantical considerations. It is therefore not *ad hoc* in the way current theories of syllogistic are, nor does it share their representational shortcomings.

4. Interpreting quantifier expressions

In the field of natural-language semantics, expressions like "all", "most", "some", etc. are analyzed as denoting relations between sets, or generalized quantifiers.[9] Thus, "All A are B" is taken to mean that the set of As is a subset of the set of Bs, while "No A are B" asserts that the intersection between the As and the Bs is empty. Formally, if we render "Q A are B" as "Q(A, B)", and use $\|X\|$ to refer to the extension of a term (i.e. the set of all Xs), "all" and "no" are interpreted as follows:[10]

$$\text{all(A, B) is true iff } \|A\| \subseteq \|B\|$$

$$\text{no(A, B) is true iff } \|A\| \cap \|B\| = \varnothing$$

This style of interpretation extends in a natural way to other quantifying expressions. For example, "Some A are B" means that the intersection between the As and the Bs is non-empty:

$$\text{some(A, B) is true iff } \|A\| \cap \|B\| \neq \varnothing$$

"Three A are B" means that the cardinality of the intersection between the As and the Bs equals three:

$$\text{three(A, B) is true iff card}(\|A\| \cap \|B\|) = 3$$

Quantifiers like "most", "many", and "few" are more challenging, because they are vague and perhaps ambiguous, to boot. This is just to say, however, that they spell trouble for *any* semantic analysis. But the general kind of meaning they convey can be captured in the present framework without further ado. For example, to a first approximation at least "Most A are B" means that the majority of the As are B, i.e. that there are more As that are B than As that aren't:

$$\text{most(A, B) is true iff card}(\|A\| \cap \|B\|) > \text{card}(\|A\| - \|B\|)$$

One of the reasons why predicate logic is inadequate as a semantics for natural language is that it cannot express this kind of meaning, which *essentially* involves reference to sets.

Viewing quantifiers as relations between sets means that we can try and capture semantic distinctions and similarities amongst quantifying expressions in terms of properties of

[9] The concept of generalized quantifier was introduced by Mostowski in 1957, and imported into natural-language semantics by Barwise and Cooper (1981), whose article remains one of the best introductions to the subject. Generalized quantifiers may be viewed not only as relations between sets, as I do here, but also as functions from sets to families of sets. From a logical point of view, one perspective is as good as the other, but the former is more natural and more adequate from a processing perspective.

[10] In these definitions I adopt the truth-conditional stance on meaning, and explicate the meaning of a sentence by specifying the circumstances under which it is true ("iff" is an abbreviation of "if and only if"). Readers not familiar with truth-conditional semantics can take "is true iff" as synonymous with "means that".

relations. There are various such properties that have proved to be especially relevant to natural-language quantification, two of which I want to single out here, viz. symmetry and monotonicity. According to the definitions just given, some quantifiers are symmetric while others are not. For example, “some”, “no”, and “three” are symmetric; “all” and “most” are not. Hence, it follows from the definitions above that the following propositions must be valid:

If some lawyers are crooks then some crooks are lawyers.
If no lawyers are crooks then no crooks are lawyers.
If three lawyers are crooks then three crooks are lawyers.

This prediction is confirmed by speakers’ intuitions. The following, on the other hand, should not be valid:

If all lawyers are crooks then all crooks are lawyers.
If most lawyers are crooks then most crooks are lawyers.

This prediction, too, appears to be correct. Non-symmetric quantifiers are universal (English “all”, “every”, and “each”) or proportional, like “most” and “half of the”. The distinction between symmetric and non-symmetric quantifiers has been shown to manifest itself in several ways, the best-known of which is that in many languages, including English, existential *there*-sentences only admit symmetric quantifiers:

$$\text{There are} \left\{ \begin{array}{c} \text{some} \\ \text{no} \\ \text{three} \\ {}^{*}\text{all} \\ {}^{*}\text{most} \end{array} \right\} \text{lawyers on the beach.}$$

The distinction between symmetric and non-symmetric quantifiers is also implicated in the interpretation of donkey sentences,[11] for example, and it plays an important role in the acquisition of quantifying expressions. It is well-known that young children tend to have difficulties interpreting propositions like “All the boys are kissing a girl”, as uttered of a scene with, say, three boys kissing one girl each plus one further girl who isn’t kissed by anyone. Children are prone to believe that the sentence is false in such a situation, but they never make analogous mistakes with symmetric quantifiers. Furthermore, it has been shown that previous exposure to sentences with symmetric quantifiers has an adverse effect on children’s performance with non-symmetric quantifiers, though not vice versa.[12] It appears, therefore, that symmetry is a key element in the acquisition of quantification, too.

Another property (or rather, family of properties) that looms large in the semantic

[11] Donkey sentences are so-called after the classic example of Geach (1962), “Every farmer who owns a donkey beats it.” See Kanazawa (1994) and Geurts (2002) for more recent discussion.

[12] Smith (1979, 1980). See Drozd (2001) and Geurts (2001) for discussion of symmetry in the context of language acquisition.

literature is monotonicity. Like symmetry, this notion is not restricted to quantifiers, and I will introduce it with the help of a non-quantified example:

Fred's tie is navy blue.
Fred's tie is blue.

Since "navy blue" entails "blue" (the latter predicate applies to everything of which the former holds), the first sentence entails the second. The position occupied by "navy blue" in the first sentence is *upward entailing* (or monotone increasing), which is to say that truth will be preserved if "navy blue" is replaced with a term it entails. Similarly, it follows from "Fred's tie isn't blue" that "Fred's tie isn't navy blue". The position occupied by "blue" in the first sentence is *downward entailing* (or monotone decreasing), which is to say that truth will be preserved if "blue" is replaced with a term it is entailed by (negation reverses monotonicity). Monotonicity is a very broad concept: in principle, any linguistic position may be upward or downward entailing, or neither (non-monotone). In particular, each quantifier has its own monotonicity profile. Consider, for example, the following proposition:

If all pachyderms are navy blue, then:
(a) all pachyderms are blue, and
(b) all elephants are navy blue.

Since everything that is navy blue is blue, (a) implies that the second argument position of "all" is upward entailing; and as "elephant" entails "pachyderm", (b) implies that the first argument position is downward entailing. Using a plus sign for upward entailing and a minus sign for downward entailing positions, we can summarize the monotonicity profile of "all" thus: all(A^-, B^+). Table 3 gives the monotonicity profiles of the syllogistic moods and two sentence schemas with cardinal quantifiers.

Note that "exactly three" is non-monotone in both of its argument positions. The following propositions, neither of which is valid, illustrate this for the first argument position:

If exactly three pachyderms are blue, then exactly three elephants are blue.
If exactly three elephants are blue, then exactly three pachyderms are blue.

The following proposition, on the other hand, *is* valid:

If three elephants are blue, then some elephants are blue.

This is because the position occupied by the quantifier "three" *itself* is upward entailing, and "three" entails "some"; it follows from the definitions given above that, for any pair of predicates A, B, if "three(A, B)" is true, then "some(A, B)" is true, as well. More generally, if we have a sentence of the form "Q(A, B)", then the position occupied by Q is upward entailing; that is to say, this property holds irrespective of the quantified expression replacing Q.

It was already mentioned in passing that negative expressions reverse monotonicity: upward becomes downward, and vice versa. For example, if "all(A^-, B^+)" occurs within the scope of a negation operator, we get "not all(A^+, B^-)", as witness the following, which

Table 3
Monotonicity profiles of some quantifiers, with diagnostic tests

	Validity test for A-*position*	*Validity test for* B-*position*
all(A^-, B^+)	If all pachyderms are pink, then all elephants are pink.	If all elephants are navy blue, then all elephants are blue.
some(A^+, B^+)	If some elephants are pink, then some pachyderms are pink.	If some elephants are navy blue, then some elephants are blue.
some(A^+, not B^-)	If some elephants are not pink, then some pachyderms are not pink.	If some elephants are not blue, then some elephants are not navy blue.
no(A^-, B^-)	If no pachyderms are pink, then no elephants are pink.	If no elephants are blue, then no elephants are navy blue.
at-least-three(A^+, B^+)	If at least three elephants are pink, then at least three pachyderms are pink.	If at least three elephants are navy blue, then at least three elephants are blue.
at-most-three(A^-, B^-)	If at most three pachyderms are pink, then at most three elephants are pink.	If at most three elephants are blue, then at most three elephants are navy blue.

are both valid:

If not all elephants are blue, then not all pachyderms are blue.
If not all elephants are blue, then not all elephants are navy blue.

There is one syllogistic mood which involves explicit negation, namely "some(A, not B)", whose monotonicity profile is: "some(A^+, not$(B^-)^+$)". Note that the position within the scope of the negation operator is downward entailing, while the argument position as such, now occupied by a negated predicate, remains upward entailing.

Monotonicity has been shown to be involved in various semantic phenomena, including donkey sentences, the semantics of temporal connectives, co-ordination, and polarity; here I will briefly illustrate the latter two. Compare the following propositions, both of which are valid:

If at least five lawyers sang *and* danced, then at least five lawyers sang and at least five lawyers danced.
If at most five lawyers sang *or* danced, then at most five lawyers sang and at most five lawyers danced.

More generally, for some Qs, we may infer from Q(A, B and C) that Q(A, B) and Q(A, C), while for other Qs, the same conclusion may be drawn from Q(A, B or C). The former pattern holds for quantifiers that are upward entailing in their second argument position, and the latter holds for quantifiers that are downward entailing in that position. Since the predicate "sing and dance" entails "sing" as well as "dance", each of which entails "sing or dance", and "at least five" and "at most five" are, respectively, upward and downward entailing in their second argument, the facts observed above follow from the monotonicity properties of the quantifiers "at least five" and "at most five".

All languages have negative polarity items, which are so-called because they typically occur within the scope of a negative expression, and are banned from positive environments. English negative polarity items are "any" and "ever", for example:

$$\text{Wilma} \left\{ \begin{array}{c} {}^{*}\text{has} \\ \text{doesn't have} \end{array} \right\} \text{any luck.}$$

$$\left\{ \begin{array}{c} {}^{*}\text{Someone} \\ \text{No one} \end{array} \right\} \text{has any luck.}$$

On closer inspection, it turns out that negative polarity items do not necessarily require a negative environment, though there certainly are constraints on where they may occur, as witness:

 If Wilma has any luck, she will pass the exam.
*If Wilma passes the exam, she must have any luck.
 Everyone who has any luck will pass the exam.
*Everyone who passes the exam must have any luck.

The generalization is that negative polarity items may only occur in downward entailing positions. In effect, a negative polarity item serves to *signal* that the environment in which it occurs is downward entailing, which goes to show that monotonicity is of some importance to languages and their speakers (Ladusaw, 1979, 1996).

The purpose of the foregoing survey was to explain why semanticists count symmetry and monotonicity among the most important properties of natural-language quantifiers. Assuming that they are right about this, it is not unreasonable to hypothesize that these properties play a role in reasoning with quantifiers, as well. I will now try to show that this hypothesis is a fertile one.

5. A monotonicity-based model of reasoning with quantifiers

In this section I present a very simple logic which builds on the observations made in the foregoing. In this logic all valid classical syllogisms are provable, but it goes far beyond traditional syllogistic logic in that it renders many other arguments valid, as well. The logic has three rules of inference, which follow directly from the interpretation of the quantifiers and negation. The logic's workhorse is monotonicity, which turns out to be implicated in every valid syllogistic argument. Once this logic is in place, it is not very difficult to produce a processing model that accounts for the data reviewed in Section 2.[13,14]

[13] I am by no means the first to observe the importance of monotonicity to syllogistic reasoning. Indeed, it may be argued that the concept is implicit in the traditional *dictum de omni* and the notion of so-called distributed occurrence of terms. The most thorough discussion of the role monotonicity plays in syllogistic inference is by Sánchez Valencia (1991).

[14] A caveat: my main concern in this paper is with the representations used in reasoning with quantifiers. The processing model presented below is my official proposal, to be sure, but whatever interest it has lies chiefly in the rules and representations it employs. I have nothing new to say about reasoning errors, and nothing at all about reasoning strategies. Concerning the latter point, I consider it quite likely that people employ different types of reasoning strategies, which may involve different types of representation (as, for example, Ford, 1995 has argued), but in this paper I confine my attention to one particular type.

To begin with, we need a formal syntax for our representation language, which is not too hard to provide, because the syntax of syllogistic logic is so simple. Matters are complicated slightly because we need a representation in which upward and downward entailing positions are made explicit, but this, too, is fairly straightforward:[15]

Vocabulary:
- basic terms: A, B, C, …
- quantifiers: all, some, no
- a special two-place predicate: $\Rightarrow$
- diacritical signs and brackets: $^+$, $^-$,), (

Syntax:
- If α is a basic term, then α^+ and $(\text{not } \alpha^-)^+$ are positive terms and α^- and $(\text{not } \alpha^+)^-$ are negative terms.
- If α is a negative term and β is a positive term, then $\text{all}^+(\alpha, \beta)$ is a sentence.
- If α and β are positive terms, then $\text{some}^+(\alpha, \beta)$ is a sentence.
- If α and β are negative terms, then $\text{no}^+(\alpha, \beta)$ is a sentence.
- If α and β are both either terms or quantifiers, then $\alpha \Rightarrow \beta$ is a sentence.

These rules generate the kind of strings we have been using already, like "$\text{all}^+(A^-, B^+)$", "$\text{some}^+(A^+, (\text{not } B^-)^+)$", and so forth. Since the position of negation is not restricted to the second term, this syntax also produces strings like "$\text{all}^+((\text{not } A^+)^-, B^+)$", for which there is no use in a syllogistic logic, but which will not be in the way, either. Other strings that aren't part of traditional logic, but are essential to ours, are of the form "$\alpha \Rightarrow \beta$", where α and β are either terms or quantifiers; this proposition may be read as "α implies β". If A and B are terms, then "A $\Rightarrow$ B" means that all As are Bs. Hence, "A $\Rightarrow$ B" and "all(A, B)" are synonymous, and will accordingly be treated as notational variants. Implication is not restricted to terms; quantifiers may imply each other, too. For example, in traditional syllogistic logic (though not in predicate logic) "all" implies "some", which is rendered in the present notation as "all $\Rightarrow$ some".

These syntactic rules define the official language of our logic. In practice, however, we will drop the brackets enclosing negated terms, as well as all diacritics save for the ones required by the occasion. Thus, whenever a diacritical plus or minus appears it flags a position that is actually used in a proof.

Our chief rule of inference is the following:

$$\frac{\alpha \Rightarrow \beta \qquad \ldots\, \alpha^+ \ldots}{\ldots\, \beta^+ \ldots} \qquad \frac{\beta \Rightarrow \alpha \qquad \ldots\, \alpha^- \ldots}{\ldots\, \beta^- \ldots} \quad \text{MON}$$

[15] For monotonicity marking in less trivial languages, see Sánchez Valencia (1991) and Dowty (1994).

In words: any expression α occurring in an upward entailing position may be replaced with any expression β that is implied by α, and any expression α occurring in a downward entailing position may be replaced with any expression β that implies α.

Our second rule of inference is based on symmetry, and its application is therefore restricted to symmetric quantifiers; it is the conversion rule used already by Aristotle:

$$\frac{Q(A, B)}{Q(B, A)} \quad \text{CONV} \qquad (Q = \text{"some" or "no"})$$

Without further provisions, MON and CONV suffice to prove 11 syllogistic arguments valid in predicate logic. In all cases the conclusion is derivable in one or two steps, using either MON alone or MON and CONV. The following proof of AE4E is as complex as it gets:

[1]	all(C, B)	*premiss*
[2]	no(B^-, A)	*premiss*
[3]	no(C, A)	MON *applied to* [1] *and* [2]
[4]	no(A, C)	CONV *applied to* [3]

Here MON applies to an argument, but the rule is not restricted to any particular category of expression, and may affect negated terms, too, as in the following proof of AO2O:

[1]	all(C, B)	*premiss*
[2]	some(A, not B^-)	*premiss*
[3]	some(A, not C)	MON *applied to* [1] *and* [2]

The remaining valid syllogisms cannot be obtained with MON and CONV alone. This is partly due to the fact that MON is as yet restricted in its application to terms, but we also need one further rule:

$$\frac{\text{no}(A, B)}{\text{all}(A, \text{not } B)} \quad \text{NO/ALL-NOT}$$

Like the conversion rule, this one follows directly from the meanings of the quantifiers involved. As it turns out, the effect of the NO/ALL-NOT rule will always be to feed into the MON rule. With our new rule, we can prove all 15 syllogisms that are valid in standard predicate logic. The following proof, of syllogism EI3O, uses all rules introduced thus far:

[1]	no(B, C)	*premiss*
[2]	some(B, A)	*premiss*
[3]	some(A, B^+)	CONV *applied to* [2]
[4]	all(B, not C)	NO/ALL-NOT *applied to* [2]
[5]	some(A, not C)	MON *applied to* [3] *and* [4]

This is a relatively long proof, but then the syllogism is not an easy one.

The remaining syllogisms are not valid in standard predicate logic, because they require the presupposition that "all" and "no" range over non-empty domains of quantification. Slightly more accurately: traditional logic has it that "all(A, B)" and "no(A, B)" entail that there are As. In terms of generalized quantifier theory, this is to say that these quantifiers are construed as follows:

$$\text{all(A, B) is true iff } \|A\| \neq \emptyset \text{ and } \|A\| \subseteq \|B\|$$

$$\text{no(A, B) is true iff } \|A\| \neq \emptyset \text{ and } \|A\| \cap \|B\| = \emptyset$$

There is a simple way of capturing this presupposition in our system, namely by adding the following axiom, which just says that "all" implies "some":

all ⇒ some ALL/SOME

Again, this addition is licensed directly by the interpretation of the quantifiers involved (as construed in traditional logic), and as with the NO/ALL-NOT rule, the main function of ALL/SOME will be to feed into the MON rule. With this new axiom, "some(A, B)" can be derived from "all(A, B)", courtesy of the MON rule, and "some(A, not B)" becomes derivable from "no(A, B)", because NO/ALL-NOT gives us "all(A, not B)", from which "some(A, not B)" follows through MON. The following proof of syllogism EA2O illustrates the use of ALL/SOME:

[1]	no(C, B^-)	*premiss*
[2]	all(A, B)	*premiss*
[3]	no(C, A)	MON *applied to* [1] *and* [2]
[4]	no(A, C)	CONV *applied to* [3]
[5]	all^+(A, not C)	NO/ALL-NOT *applied to* [4]
[6]	some(A, not C)	MON *applied to* [5] *and* ALL/SOME

Thus, all valid arguments can be accounted for with a handful of inference rules that follow directly from the semantics of the logical vocabulary of syllogistic logic: "all", "some", "no", and "not".

What remains to be shown is how this logic can be embedded in a processing model. In principle, there are many ways of doing this, but for current purposes it will suffice to show that even a crude processing model can produce reasonable predictions. Let us assume, therefore, that inference rules are applied in a breadth-first fashion until the right sort of conclusion is found or no new inferences can be made. What the "right sort of conclusion" is depends on the task. In an evaluation paradigm, it is the conclusion specified by the experimenter, or its negation; in a multiple-choice paradigm, any one of the given conclusions is of the right sort; and in a production paradigm, any sentence of the syllogistic language is of the right sort.[16] Since inference rules are applied breadth-first, the system is guaranteed to find a minimal proof that isn't longer than any other proof (if a proof exists, that is). In many cases, there will be more than one minimal proof of a valid syllogism, but these will only differ in the order in which inference steps are made: the rules will be the same, and so will the number of inferences.[17]

As is common in logic-based accounts, I take it that the complexity of a syllogism is determined chiefly by the number of inference steps needed to get from the premisses to the conclusion. In the present case, this is to say that the length of any minimal proof is the main predictor. But there is another factor, as well, viz. grammatical structure. It is a well-established fact that more syntactic structure makes a sentence harder to process, and as deduction tasks always involve sentence processing, it doesn't come as a surprise that grammatical complexity plays a role in reasoning, too. Grammatically speaking, three quarters of all syllogistic propositions have the same structure: "Q A are B". However, O-propositions have the form "Some A are not B", and should therefore be harder to process than propositions in the other moods.

Putting these considerations together, I propose the following model. Our abstract reasoner starts out with a budget of 100 units, which are used to pay for inferences and grammatical complexity, according to the following rules:[18]

- For every use of MON, subtract 20 units.
- For every use of NO/ALL-NOT, subtract 10 units.
- If a proof contains an O-proposition, subtract 10 units.

For reasons discussed in Section 2, I assume that CONV is for free. That the NO/ALL-NOT rule is cheaper than MON is plausible, too, because the latter rule combines information from two propositions, whilst the former merely maps one proposition onto another. Table 4 shows the predicted difficulty of all valid syllogisms alongside the scores of Chater and Oaksford's meta-study (cf. Table 1). The correlation between the two is good ($r = 0.93$).

We now have a monotonicity-based model which accounts quite well for people's performance on valid syllogisms, which was one of our main objectives, because validity is the major factor in syllogistic reasoning, as I argued in Section 2. In the same section, we saw that many errors in syllogistic reasoning can be put down to illicit conversion of propositions with "all" and "some … not". This is something that is easily incorporated in our model. We only need to extend CONV so that it applies not only to propositions with

[16] More sophisticated models can be obtained by refining the notion of "right sort of conclusion", which is somewhat simplistic as it stands. Such refinements should account for the fact that we prefer to draw conclusions that are non-trivial and relevant to our current purposes – which may be rather a tall order.

[17] As the number of valid syllogisms is quite small, this can easily be proved by enumeration of alternatives.

[18] Of course, this talk of "reasoning budgets" is merely a picturesque alternative to the common procedure of assigning numerical weights to inference rules. It must be admitted that it is not entirely clear what such weights stand for. The basic idea surely is that weights represent processing effort, but this notion is inappropriate if we allow for illicit inference rules. I will not attempt to sort out this matter here.

Table 4
Predicted difficulty of valid syllogisms according to the model described in the text, compared with Chater and Oaksford's scores (in parentheses)

AA1A	80	(90)	OA3O	70	(69)	EA1O	40	(3)
EA1E	80	(87)	AO2O	70	(67)	EA2O	40	(3)
EA2E	80	(89)	EI1O	60	(66)	EA3O	40	(22)
AE2E	80	(88)	EI2O	60	(52)	EA4O	40	(8)
AE4E	80	(87)	EI3O	60	(48)	AE2O	40	(1)
IA3I	80	(85)	EI4O	60	(27)	AE4O	40	(2)
IA4I	80	(91)	AA1I	60	(5)			
AI1I	80	(92)	AA3I	60	(29)			
AI3I	80	(89)	AA4I	60	(16)			

"some" and "no" but also to propositions with "all" or "some … not". However, we still want to differentiate between licit and illicit conversion, because the latter is less common than the former. Therefore, we assume that, unlike its legal counterpart, illicit conversion is not for free: it costs 20 units. Even with illicit conversion, most syllogisms remain unprovable, and we simply assume that an unprovable syllogism sets the reasoner back by 80 units, which is the price of the most difficult argument that does have a proof (with illicit conversion).[19] This model makes quite reasonable predictions for the complete set of syllogisms, with $r = 0.83$, and if we set aside the syllogisms which are probably undervalued by Chater and Oaksford's figures because, in the experiments analyzed by Chater and Oaksford, they had to compete with other syllogisms, then $r = 0.88$.

The main virtue that I claim for my account is that it extends in a natural way beyond the confines of traditional syllogistic logic. For example, it is a trivial exercise to incorporate cardinal quantifiers, like "at least *n*". From a semantical point of view, "at least *n*" is of the same type as "some": both are symmetric quantifiers that are upward entailing in both of their argument positions. The proposed account predicts, therefore, that arguments with "at least *n*" will be equally complex as corresponding arguments with "some", regardless the size of *n*.

Ceteris paribus, I would predict that "at most *n*" affects the complexity of an argument in the same measure as "at least *n*" does, for the following reason. The main difference between "some" and "no" is that whereas the former is upward entailing the latter is downward entailing in both of its argument positions. Therefore, whenever we have commensurable arguments with "some" and "no", they should be equally complex. This prediction is borne out by the data (see the Chater and Oaksford (1999) figures for AEE/EAE and AII/IAI arguments). Moreover, "at most *n*" is of the same semantic type as "no": they are both symmetric quantifiers that are downward entailing in both argument positions. Hence, by transitivity, "at least *n*", and "at most *n*" should be equally difficult.

However, all things are not equal: considerations extraneous to the proposed model

[19] This is admittedly stipulative, but it is not entirely arbitrary because it means, in the present model, that the reasoning system begins to falter after four or five inference steps – which seems quite reasonable to me. Still, this is a matter that calls for a more refined treatment.

suggest that "at most *n*" may be more difficult than "at least *n*". There is a wealth of linguistic and psychological evidence which shows that in pairs like "tall–short", "many–few", "happy–unhappy", etc., the first member, which is in a sense the positive one, enjoys a privileged status (see Horn, 1989 for a survey). Linguistically, the negative form is marked, which means that it does not figure in all environments that admit its positive counterpart. For example, one normally would ask, "How tall is Fred?", not "How short is he?". Psychologically, negative expressions take longer to process, cause more errors, and are harder to retain than positive ones. Now, it seems likely that "at least *n*–at most *n*" will follow the pattern of "tall–short", "many–few", and "happy–unhappy", and if it does, arguments with "at most *n*" will be more difficult than arguments with "at least *n*", presumably because the representation of "at most *n*" contains a negative element: "At most *n* A are B" is represented as "Not more than *n* A are B". In terms of our semantical framework, this means that we must not interpret "at most *n*" directly:

$$\text{at-most-}n(\text{A, B}) \text{ is true iff card}(\|\text{A} \cap \text{B}\|) \leq n$$

Instead, "at most *n*" is to interpreted as the negation of "more than *n*":

$$\text{more-than-}n(\text{A, B}) \text{ is true iff card}(\|\text{A} \cap \text{B}\|) > n$$

From a logical point of view, these interpretations are equivalent ("at-most-*n*(A, B)" and "not more-than-*n*(A, B)" always have the same truth value), but linguistically as well as psychologically they are different.

To summarize: I predict that "at least *n*" is of the same complexity level as "some", for any *n*, whereas "at most *n*" is more difficult. In order to test these predictions, I conducted an experiment in which subjects were presented with syllogistic arguments involving (the Dutch equivalents of) "some", "at least *n*", and "at most *n*", where *n* was an integer between 20 and 30 (the variation was used as a precaution against interference between tasks). The terms of each syllogism were randomly selected from a small collection of nouns like "forester", "communist", "poet", and so on. For each quantifier Q, there were four arguments to be assessed:

Figure 1	*Figure 2*	*Figure 3*	*Figure 4*
All B are C	All C are B	All B are C	All C are B
Q A are B	Q A are B	Q B are A	Q B are A
Q A are C	Q A are C	Q A are C	Q A are C

Note that the arguments in figures 1 and 3 are valid if the B-positions in "Q A are B" and "Q B are A" are upward entailing, and invalid otherwise; similarly, the arguments in figures 2 and 4 are valid if the B-positions in "Q A are B" and "Q B are A" are downward entailing, and invalid otherwise. With three quantifiers and four argument schemata, there were 12 syllogistic arguments altogether, which were alternated with one-premiss arguments like the following:

Table 5
Percentage of correct responses in the experiment described in the text, with standard deviations in parentheses

	1 premiss	*2 premisses*	*Valid*	*Invalid*	*All*
At least	97 (9)	92 (14)	96 (10)	93 (14)	95 (8)
Some	95 (11)	90 (12)	93 (14)	91 (14)	92 (8)
At most	89 (15)	67 (24)	70 (21)	87 (22)	78 (15)

At least 24 communists own a blue bicycle.

At least 24 communists own a bicycle.

Note that this is a monotonicity argument, too, though it should be easier than the corresponding figure 1 syllogism, because it is shorter.

Since I had to make do without the usual experimental facilities, I cajoled 23 friends and relations into taking the test. All participants were native speakers of Dutch with an academic degree in psychology or linguistics, but no previous exposure to logic.

The results of the experiment are presented in Table 5.[20] To analyze these data, a repeated measures ANOVA was conducted with three within-subject factors: quantifier ("at least", "at most", "some"), argument length (one or two premisses), and validity (valid or invalid). This yielded main effects for quantifier ($F(2, 44) = 14.533$, $P < 0.001$) and argument length ($F(1, 22) = 12.517$, $P < 0.002$), but not for validity. There were interactions between quantifier and argument length ($F(2, 44) = 6.466$, $P < 0.009$) and quantifier and validity ($F(2, 44) = 4.926$, $P < 0.018$). Further analysis of these two interactive effects tied them to arguments featuring "at most"; in both cases there were significant differences between arguments with "at most" and "some" (quantifier/argument length: $P < 0.010$; quantifier/validity: $P < 0.033$) and between arguments with "at most" and "at least" (quantifier/argument length: $P < 0.017$; quantifier/validity: $P < 0.016$). There were no significant differences between "at least" and "some". In order to determine if any of the differences between arguments with the same quantifier were significant, *t*-tests were conducted with quantifier and argument length and quantifier and validity as factors. These tests, too, attained significance only for arguments with "at most": $t = 3.792$ ($P < 0.001$, two-tailed) and $t = -2.577$ ($P < 0.017$, two-tailed), respectively.

These results are consistent with our main predictions: that there is no relevant difference between "some" and "at least", and that arguments with "at most" are more difficult. But at the same time they cloud the picture somewhat, because it turns out that the strictly additive measure of complexity that underlies our model is not quite adequate. It is not as if *any* argument with "at most" is harder than parallel arguments with "some" or "at least"; rather, it is valid and/or two-premiss arguments with "at most" that are more difficult than others. This, however, is a concern not only for the present proposal but for all current theories of deductive reasoning.

[20] I am indebted to Frans van der Slik for carrying out the analyses reported in the following and helping me interpret the results.

Of the two interactions found in this study, the one between quantifier type and validity is the most troubling, in my view. Earlier on in this paper I argued that valid arguments tend to be easier than invalid ones (see Section 2), and now we find that some valid arguments are harder than their invalid counterparts. This need not be a contradiction, of course, but I do believe that there is a serious problem lurking here. It is that thus far we lack a good understanding of why people reject some arguments as “not valid” or maintain that “nothing follows” from a given set of premisses. If someone says that a conclusion φ does not follow, it may be either because he has a proof of “not φ” or because he doesn’t know how to prove φ. These are quite different things, obviously, but the evaluation task used in our experiment doesn’t distinguish between the two. Other experimental techniques are more discriminating in this respect, but even the paradigms which allow subjects to say that “nothing follows” are relatively crude instruments because there is likely to be more than one possible reason why someone should think that “nothing follows”; for example, he may judge that a given conclusion, though correct, is pointless or odd.[21] In brief, this is a topic that calls for more, and better, experimentation.

6. Concluding remarks

One popular way of characterizing logical inference is that a conclusion φ follows logically from a set of premisses $\psi_1 \ldots \psi_n$ if the *meanings* of φ and $\psi_1 \ldots \psi_n$ alone guarantee that φ is true if $\psi_1 \ldots \psi_n$ are. It is not the facts but the meanings of its component propositions that render an argument valid or invalid. Hence, in order to understand logical inference we must understand how arguments are interpreted: no inference without interpretation. I have endeavoured to demonstrate that this slogan applies with a vengeance to syllogistic reasoning.

The main virtues of the model I have presented are the following. First and most importantly, my account is based on a system of inference that is independently motivated by the meaning of its logical vocabulary: “all”, “no”, “some”, and “not”. Secondly and relatedly, this system can be extended in a straightforward and principled way not only to the non-classical quantifiers but across the board. Thirdly, the model predicts a complexity

[21] A case in point is the well-known fact that the seemingly trivial step from “It is raining” to “It is raining or snowing” is actually quite hard to take, though it doesn’t seem right to say that the inference is especially complex; it is just *odd* that someone should want to draw this conclusion. Some researchers have, implicitly or explicitly, rejected this diagnosis. Thus, Braine, Reiser, and Rumain (1984) set up their “mental logic” in such a way that it is very hard to derive “φ or ψ” from φ alone. However, this also makes the following argument virtually impossible to prove:

$$\frac{\begin{array}{l}\varphi \\ \text{If } \varphi \text{ or } \psi\text{, then } \chi\end{array}}{\chi}$$

Subjects typically find it very easy to see that this is valid, and therefore Braine et al. have no choice but to stipulate that this is a valid pattern of inference. I have criticized such manoeuvres in Section 3, and argued that they should be avoided at all costs. There is quite a bit more to say about this matter, but I will not say it here.

ranking that fits well with the experimental data. Fourthly, the current proposal is simpler than any other theory that covers the same ground, including "fast and frugal" heuristic models of syllogistic reasoning like Chater and Oaksford's.

Methodological considerations aside, the key element in my proposal, which distinguishes it from all previous accounts in the psychological literature, is that it drops the assumption that syllogistic reasoning is always in terms of *individuals*. Generalized-quantifier theory leads us to expect that reasoning with quantifiers is done in terms of *sets* instead, and I have tried to show that a processing model based on this assumption can be quite successful.

Logic-based approaches to deduction have been criticized on a number of counts. There is a popular view that ordinary folk are bad at logical reasoning, and that, consequently, it is a priori unlikely that they employ anything like a mental logic. A related argument, advanced by Chater and Oaksford (Chater & Oaksford, 1999; Oaksford & Chater, 2001), among others, is that everyday reasoning is not logical, so that whatever it is people do when they solve deduction tasks cannot be logic. Arguments along these lines invariably rely on carefully selected evidence. To a large extent, the rumour that people aren't good at logic is based on experimental data on conditional reasoning. In particular, it has been demonstrated again and again that subjects fail in large numbers on certain versions of the Wason task. But then conditionals rank high among the more controversial topics in semantics and the philosophy of language; at present, it is simply unclear what their logic *is*, and therefore we lack a sound normative theory against which subjects' performance can be assessed. Moreover, even if it had been established that performance on *some* conditional-reasoning tasks is poor from a logical point of view, there are scores of logical inferences that people are quite good at, like the following, for example:

The butler and the chauffeur have an alibi.
___
The chauffeur has an alibi.

I take it to be self-evident that very few people will have problems with this, and the experimental work of Braine, Reiser, and Rumain (1984) proves, if proof is required, that there are lots of arguments like this. Such bread-and-butter inferences tend to pass unnoticed, but we are making them all the time, and it would be far-fetched to deny that they are logical inferences, pure and simple.

Another objection against logic-based accounts of reasoning has been made by evolutionary psychologists (e.g. Cosmides, 1989; Cosmides & Tooby, 1992; Gigerenzer & Hug, 1992). Logic-based theories, so the criticism goes, assume that we have a mental logic which is a domain-independent instrument for reasoning about anything that happens to arouse our interest. But how did we acquire this general-purpose tool? There is precious little evidence that our parents teach us how to reason, nor is it clear how evolution could have hit upon such a device. It seems rather likely that Mother Nature equipped us with specialized modules for reasoning about physical objects, social relationships, snakes and spiders, and so forth, but it is utterly mysterious how a full-blown logical faculty could be the outcome of natural selection. I sympathize with this line of argument, and have used it myself to criticize previous logic-based theories of reasoning (Section 3). However, this criticism takes for granted a view on mental logic that I believe is wrong. Though it may be that some elementary notions of logic are innate, mental logic must not be conceived of as an autonomous module of the mind. As I have said a number of times already, and as I have illustrated in the foregoing with syllogistic reasoning, there has to be an intimate connection between the meaning of an expression and valid arguments which make essential use of that expression. Mental logic is largely a concomitant of our linguistic prowess, and though it is still a matter of controversy where *that* came from, nobody will doubt that we *have* it.

Acknowledgements

I am greatly indebted to Rob van der Sandt and three anonymous readers of *Cognition* for their elaborate and very constructive comments on earlier versions of this paper, and to Frans van der Slik for his statistical advice and for carrying out the analyses reported in Section 5.

References

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, *4*, 159–219.

Braine, M. D. S. (1998). Steps towards a mental-predicate logic. In M. D. S. Braine & D. P. O'Brien (Eds.), *Mental logic* (pp. 273–331). Mahwah, NJ: Lawrence Erlbaum Associates.

Braine, M. D. S., Reiser, B. J., & Rumain, B. (1984). Some empirical justification for a theory of natural propositional logic. In G. H. Bower (Ed.), *The psychology of learning and motivation*. New York: Academic Press.

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

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

Cosmides, L. (1989). The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*, 187–276.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: evolutionary psychology and the generation of culture* (pp. 163–228). Oxford: Oxford University Press.

Dickstein, L. S. (1978). The effect of figure on syllogistic reasoning. *Memory and Cognition*, *6*, 76–83.

Dickstein, L. S. (1981). Conversion and possibility in syllogistic reasoning. *Bulletin of the Psychonomic Society*, *18*, 229–232.

Dowty, D. (1994). The role of negative polarity and concord marking in natural language reasoning. In M. Harvey & L. Santelmann (Eds.), *Proceedings from Semantics and Linguistic Theory IV* (pp. 114–144). Ithaca, NY: Cornell University Press.

Drozd, K. F. (2001). Children's weak interpretations of universally quantified questions. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 340–376). Cambridge University Press: Cambridge.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: the psychology of deduction*, Hove: Lawrence Erlbaum.

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

Geach, P. T. (1962). *Reference and generality*. Ithaca, NY: Cornell University Press.

Geurts, B. (1999). *Presuppositions and pronouns*. Oxford: Elsevier.

Geurts, B. (2001). *Quantifying kids*. Unpublished manuscript.
Geurts, B. (2002). Donkey business. *Linguistics and Philosophy*, *25*, 129–156.
Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: social contracts, cheating and perspective change. *Cognition*, *43*, 127–171.
Horn, L. R. (1989). *A natural history of negation*. Chicago, IL: Chicago University Press.
Johnson-Laird, P. N. (1983). *Mental models*, Cambridge: Cambridge University Press.
Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition*, *16*, 1–61.
Johnson-Laird, P. N., Byrne, R. M. J., & Tabossi, P. (1989). Reasoning by model: the case of multiple quantification. *Psychological Review*, *96*, 658–673.
Johnson-Laird, P. N., & Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology*, *10*, 64–99.
Kanazawa, M. (1994). Weak vs. strong readings of donkey sentences and monotonicity inference in a dynamic setting. *Linguistics and Philosophy*, *17*, 109–158.
Ladusaw, W. A. (1979). *Polarity sensitivity as inherent scope relations*. PhD dissertation, University of Texas at Austin, Austin, TX.
Ladusaw, W. A. (1996). Negation and polarity items. In S. Lappin (Ed.), *The handbook of contemporary semantic theory* (pp. 321–341). Oxford: Blackwell.
Newstead, S. E., & Griggs, R. A. (1983). Drawing inferences from quantified statements: a study of the square of oppositions. *Journal of Verbal Learning and Verbal Behavior*, *22*, 535–546.
Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, *5*, 349–357.
Rips, L. J. (1994). *The psychology of proof: deductive reasoning in human thinking*. Cambridge, MA: MIT Press.
Sánchez Valencia, V. M. (1991). *Studies on natural logic and categorial grammar*. Doctoral dissertation, University of Amsterdam, Amsterdam.
Smith, C. L. (1979). Children's understanding of natural language hierarchies. *Journal of Experimental Child Psychology*, *27*, 437–458.
Smith, C. L. (1980). Quantifiers and question answering in young children. *Journal of Experimental Child Psychology*, *30*, 191–205.
Sundholm, G. (1983). Systems of deduction. In D. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (pp. 133–188), Vol. 1. Dordrecht: Reidel.
Tsal, Y. (1977). Symmetry and transitivity assumptions about a nonspecified logical relation. *Quarterly Journal of Experimental Psychology*, *29*, 677–684.
Wetherick, N. E., & Gilhooly, K. J. (1990). Syllogistic reasoning: effects of premiss order. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie & G. Erdos (Eds.), *Lines of thinking: reflections on the psychology of thought* (pp. 99–108), Vol. 1. Chichester: Wiley.
Wilkins, M. C. (1928). The effect of changed material on ability to do formal syllogistic reasoning. *Archives of Psychology*, *16*, 1–83.