Cognitive Psychology **38,** 191–258 (1999)
Article ID cogp.1998.0696, available online at http://www.idealibrary.com on IDEAL®

The Probability Heuristics Model of Syllogistic Reasoning

Nick Chater

University of Warwick, Coventry, United Kingdom

and

Mike Oaksford

Cardiff University, Wales, United Kingdom

A probability heuristic model (PHM) for syllogistic reasoning is proposed. An informational ordering over quantified statements suggests simple probability based heuristics for syllogistic reasoning. The most important is the ''*min*-heuristic'': choose the type of the least informative premise as the type of the conclusion. The rationality of this heuristic is confirmed by an analysis of the probabilistic validity of syllogistic reasoning which treats logical inference as a limiting case of probabilistic inference. A meta-analysis of past experiments reveals close fits with PHM. PHM also compares favorably with alternative accounts, including mental logics, mental models, and deduction as verbal reasoning. Crucially, PHM extends naturally to generalized quantifiers, such as *Most* and *Few,* which have not been characterized logically and are, consequently, beyond the scope of current mental logic and mental model theories. Two experiments confirm the novel predictions of PHM when generalized quantifiers are used in syllogistic arguments. PHM suggests that syllogistic reasoning performance may be determined by simple but rational informational strategies justified by probability theory rather than by logic.

We thank Becki Grainger for running the experiments. We thank Gerd Gigerenzer and a number of anonymous reviewers who have commented on aspects of this research. We especially thank John R. Anderson who has commented extensively on many drafts of this paper and whose input has both deepened the theory and sharpened the presentation. We also thank the Cognitive Science Reading Group at Warwick University and Alan Allport, Jonathan Evans, Phil Johnson-Laird, and Ken Manktelow for valuable discussion of this work. We gratefully acknowledge a grant awarded by the Leverhulme Trust which supported Becki Grainger while she ran these experiments.

Address correspondence and reprint requests to Nick Chater, Department of Psychology, University of Warwick, Coventry, CV4 7AL, UK, or to Mike Oaksford, School of Psychology, Cardiff University, PO Box 901, Cardiff, CF1 3YG, Wales, UK. E-mail: nick.chater@warwick.ac.uk or oaksford@cardiff.ac.uk, who will also provide full mathematical derivations of the notion of ''probabilistic validity'' employed in this article.

0010-0285/99 $30.00

Almost all aspects of cognition, from perception to problem solving, involve transforming given information into a new and more useful form. It is typically assumed that this transformation can be viewed as a process of reasoning—given information can be viewed as premises, the new information as a conclusion, and the transformation can be justified in terms of a theory of inference. A key question for cognitive psychology concerns the appropriate theory of inference and how it is realized in the cognitive system. Although important throughout cognition, these questions have been most directly addressed in experimental research on the psychology of reasoning, where people are explicitly given premises and asked to draw conclusions from them.

Reasoning theorists have typically assumed that the appropriate theory of inference is provided by formal logic. Most of these accounts agree that logic provides a ''computational level'' (Marr, 1982) or ''competence'' theory of reasoning, i.e., the theory of *what* inferences people *should* draw. Controversy has centered on the cognitive processes that explain *how* people carry out those inferences, i.e., what is the appropriate algorithmic level analysis (Marr, 1982) or ''performance'' theory of reasoning. The two most well-known algorithmic accounts are mental logic (e.g., Braine, 1978; Rips, 1994) and mental models (e.g., Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991). We have argued that, if these accounts in the psychology of reasoning can provide insights into the inferential processes required in other areas of cognition, then they must generalize to commonsense, everyday reasoning (Oaksford & Chater, 1992, 1993, 1995b). That is, these theories must apply to the normal patterns of inference seen outside the reasoning laboratory. Research in a variety of areas has suggested that everyday inference is uncertain or defeasible (e.g., in artificial intelligence, McCarthy & Hayes, 1969; in cognitive psychology/cognitive science, Holland, Holyoak, Nisbett, & Thagard, 1986; Holyoak & Spellman, 1993; in philosophy and semantics, Barwise & Perry, 1983; Fodor, 1983). For example, from *Tweety is a bird* and *birds fly*, we may infer that *Tweety can fly*. But this conclusion is uncertain, because it may be defeated when we learn that Tweety is an ostrich. Such defeasible or *nonmonotonic* reasoning is beyond the scope of standard logic, which provides the competence theory for psychological accounts such as mental logic and mental models (Chater & Oaksford, 1990, 1993; Cherniak, 1986; Fodor, 1983; Oaksford & Chater, 1991, 1992, 1993, 1995b, 1998a).[1] Thus, there appears to be a theoretical mismatch between accounts in the psychology of reasoning, which are based on logic, and patterns of

[1] Some logicians view monotonicity as part of the definition of a logic, because nonmonotonicity violates the standard notion of logical validity—the conclusion can be false even if the premises are true (Curry, 1956; Curry & Feys, 1958). Nonetheless, nonmonotonic logics have been developed in an attempt to capture the defeasible pattern of human reasoning (Lewis, 1973; Reiter, 1980, 1985; Stalnaker, 1968), but these have not been applied in psychology.

everyday reasoning, which cannot be based on logic. This theoretical mismatch is mirrored by an empirical mismatch between logic and people's reasoning performance on laboratory reasoning tasks. Most psychological theories attempt to explain the mismatch between a presumed logical competence and non-logical performance, at the algorithmic level by appeal to cognitive limitations and the heuristics used to implement logic in the mind.

Recently, we have argued that these mismatches can be resolved by inverting the research strategy in the psychology of reasoning. Rather than starting from the goal of modeling performance on logical reasoning tasks and attempting to generalize to everyday inference, we start from attempting to model everyday inference, and attempt to generalize to laboratory reasoning performance. This approach is part of a more general project of the *rational analysis* of human reasoning (Oaksford & Chater, 1994, 1996, 1998a), which involves trying to explain cognitive performance in the laboratory in terms of the adaptive function of cognitive mechanisms in the real world (Anderson, 1990, 1991a, 1991b). Specifically, following proposals in artificial intelligence (e.g., Pearl, 1988), we have proposed that the appropriate computational level or competence theory for defeasible everyday reasoning is not logic, the calculus of certain reasoning, but probability theory, the calculus of uncertain reasoning (Oaksford & Chater, 1994, 1995b, 1996, 1998a). That is, instead of explaining people's nonlogical performance in laboratory tasks at the algorithmic level, we explain it at the computational level: People may not be trying and failing to do logical inference, but rather succeeding in applying probabilistic reasoning strategies.

This strategy has been applied to one of the classic laboratory reasoning tasks, Wason's selection task (1966, 1968), which reveals a stark contrast between people's performance and logical expectations (Oaksford & Chater, 1994, 1996). Oaksford and Chater (1994) assumed that people were attempting to maximize the probability of gaining information about the truth of a hypothesis rather than doing any kind of logical inference. They showed that the predictions of this probabilistic competence theory correspond closely with observed performance across a wide range of variants of the selection task. Thus, rather than viewing people as failed logicians, they suggested that people may be successful probabilistic reasoners. This result overturns the standard interpretation of the selection task, that people's natural reasoning strategies are irrational. In Oaksford and Chater's view, people's performance on the selection task is a natural and rational application of probabilistic reasoning strategies which are adapted to reasoning in the everyday world. Thus, the mismatch between logic and performance is resolved not by invoking flawed logical algorithms, but by adopting a probabilistic rather than a logical computational level theory (see also Thagard, 1988). Further research developing this viewpoint has either shown results that are consistent with this approach (Kirby, 1994; Manktelow, Sutherland, & Over, 1995; Oaksford & Chater, 1995; Oaksford, Chater, Grainger, & Larkin, 1997) or has developed theoretical variants of the approach (Evans & Over, 1996a, 1996b; Klauer, in press; Nickerson, 1996; Oaksford & Chater, 1996, 1998b).

Although encouraging for a probabilistic account of reasoning, it may be thought that this apparent success would not carry over to core logical reasoning tasks. This viewpoint is perhaps supported by the fact that some proponents of logical approaches have argued that participants do not interpret the selection task as a logical reasoning task (e.g., Rips, 1990). This suggests that these probabilistic and logical approaches may apply to disjoint sets of phenomena and, hence, do not stand in genuine competition—the selection task has simply been misclassified as a logical reasoning task. Consequently, in order to assess the general hypothesis that human everyday reasoning is uncertain and, hence, that probabilistic reasoning strategies will be imported into the laboratory, we must look to laboratory tasks which are regarded as unequivocally logical. If the probabilistic approach is to provide a general account of human reasoning, then it should also explain performance in core logical reasoning tasks.

Aside from the selection task, perhaps the most intensively researched and theoretically important task in the study of logical reasoning is syllogistic reasoning (Dickstein, 1978; Ford, 1994; Guyote & Sternberg, 1981; Johnson-Laird, 1983; Newstead, 1989; Newstead & Griggs, 1983; Polk & Newell, 1995; Rips, 1994; Stenning & Oberlander, 1995; Stenning & Yule, 1997; Woodworth & Sells, 1935). Syllogisms involve two quantified premises, each of which can have the forms All X are Y, Some X are Y, No X are Y, or Some X are not Y, where X is the subject term, and Y is the predicate term.

$$\begin{array}{l} \text{All Beekeepers are Artists,} \\ \underline{\text{All Chemists are Beekeepers}} \\ \therefore \text{ All Chemists are Artists.} \end{array}$$

In this argument, ''Beekeepers'' forms the ''middle term,'' which occurs in both premises, and the conclusion relates the ''end terms,'' ''Artists'' and ''Chemists,'' which occur in only one premise.

The main empirical finding is that performance is good on the logically valid syllogisms, but people systematically and erroneously assert conclusions for syllogisms with no valid conclusion. Syllogisms are unequivocally logical—indeed, from Aristotle until the beginnings of modern logic in the nineteenth century, syllogisms formed almost the whole of logic. Moreover, the major logic-based theories of reasoning either were first applied to syllogistic reasoning (mental models, Johnson-Laird, 1983) or have treated explaining syllogistic reasoning as a major goal (mental logic, Rips, 1994). Syllogistic reasoning, therefore, provides an ideal case study for the probabilistic approach. It is unequivocally logical; there are strong theoretical alternatives against which a probabilistic account can be compared; and there is

X - Y	Y - X	X - Y	Y - X
Y - Z	Z - Y	Z - Y	Y - Z
(4)	(1)	(2)	(3)

FIG. 1. Syllogistic figures.

a large body of empirical data, against which these theoretical proposals can be evaluated.

In this paper, we develop computational and algorithmic level analyses of syllogistic reasoning, based on a probabilistic approach. We will call the resulting model the Probability Heuristics Model (PHM). This paper has four sections. First, we outline PHM. Second we evaluate PHM in a meta-analysis of existing data. Third, we compare PHM against alternative theories. Fourth, we experimentally test PHM's novel empirical predictions. An important consequence of PHM is that it extends directly to syllogisms involving the generalized quantifiers *most* and *few* (Barwise & Cooper, 1981; Moxey & Sanford, 1991). The crucial feature of these syllogisms is that they can not be explained logically and, hence, they fall outside the scope of theories like mental logic and mental models (although see Johnson-Laird, 1983) that assume standard logic as their computational level theory. Our new experiments introduce generalized quantifiers into syllogistic arguments and, hence, provide a critical test of PHM.

THE PROBABILITY HEURISTICS MODEL

We first introduce the terminology associated with syllogisms. Syllogisms involve two quantified premises. These are: All X are Y (A), Some X are Y (I), No X are Y (E), and Some X are not Y (O) (the letters in parentheses are the traditional labels for each quantifier). We deal with two further premise types, *Most* (M) and *Few* (F). These premises can occur in one of four ''figures'' shown in Fig. 1.

Figure 1 shows the traditional figure numbering in parentheses.[2] Syllogis-

[2] The definition and numbering of the figures can cause confusion. For Aristotle, the figure is driven by the form of the conclusion, e.g., in figure 1 (see Fig. 2), the only permissible conclusion must be in the fixed order Z–X. This is because, for Aristotle, the end term of the second premise (Z) must be the subject term of the conclusion and the end term of the first premise (X) must be the predicate term of the conclusion. Confusion can arise when comparing experiments that have used this classical definition of figure with Johnson-Laird's definition of figure that depends solely on the premises and allows conclusions in free order (Z–X or X–Z). We illustrate this by an error in Evans, Newstead, and Byrne (1993), who say, ''In figure 1 syllogisms, the information is presented in the order A–B, B–C (*or B–C, A–B in the traditional task*) and participants are required to draw the conclusion A–C.'' Notice that, in the traditional task, the only allowable conclusion from A–B, B–C, is C–A, *not* A–C. Moreover, *A–B, B–C therefore C–A* is *not* a figure 1 syllogism; it is the classical figure 4 syllogism. The only possible identification of classical figure with the notion of figure used by Johnson-

tic conclusions relate the end terms, X and Z, of the premises. There are traditionally (i.e., using just AIEO premises) 64 possible forms of syllogistic premises (four quantifier combinations for the each of the two premises × four figures). Each syllogism is identified by the letters for the quantifiers it contains and its figure, e.g., AE4: first premise: *All X are Y,* second premise: *No Y are Z.* Using the standard logical interpretation of the quantifiers, there are 22 syllogisms with logically valid conclusions. Aristotle also assumed that the statements *All X are Y* and *No X are Y* presuppose that there are some Xs, which allows five more valid syllogisms (see Table 1). We denote conclusions which are valid by making such an ''existential presupposition'' by ''∃p.''

Having introduced the necessary terminology, we now outline the computational and algorithmic level theories that together make up the Probability Heuristics Model of syllogistic reasoning. We begin by outlining the algorithmic level heuristics from which our empirical predictions are derived. We then show how these provide a fast and frugal way of drawing probabilistically valid (*p*-valid; see below) inferences via a probabilistic computational level analysis.

The Algorithmic Level

The algorithmic level analysis consists of a set of ''fast and frugal'' heuristics (Gigerenzer & Goldstein, 1996), which *generate* likely syllogistic conclusions. People may also employ test processes for assessing whether generated conclusions are valid. However, we assume that, in most people, these are not well developed, which enables us to explain why many people frequently produce conclusions which are not logically valid. Moreover, these test procedures might very well involve the kinds of processes discussed by other theories that can account for logical performance, such as mental models (e.g., Johnson-Laird & Byrne, 1991), mental logic (Rips, 1994), or reasoning with Euler circles (Stenning & Oberlander, 1995). In our account, we focus on generation heuristics and assume only two simple test heuristics are used across participants. These heuristic processes of generation closely approximate the prescriptions of a probabilistic computational level analysis (see below) while imposing minimal computational overheads—they provide ''fast and frugal'' heuristics, in the sense of Gigerenzer and Goldstein (1996). We will see below that these simple heuristics account for a great deal of the data without the need to appeal to the elaborate test processes which are the focus of other accounts.

All these heuristics rely on an ordering in the informativeness of quantified statements that serve as premises of syllogistic arguments. Intuitively infor-

---

Laird, is as we have shown in Fig. 2, where X–Y, Y–Z is identified as figure 4 and Y–X, Z–Y as figure 1. Note that this is not a problem for Johnson-Laird. When reporting the effects of figure, Johnson-Laird does not refer to them using the classical numbering.

TABLE 1
The Valid Syllogisms Showing Those Conclusions That Are Valid According to (i) Aristotle, Where the Conclusion Type Is in a Fixed Order and ∃p is Allowed; (ii) Johnson-Laird, Where the Conclusion Type Is in Free Order and ∃p Is Allowed; and (iii) First Order Predicate Logic Where the Conclusion Type Is in Free Order but ∃p Is Not Allowed

Syllogism	(i) Aristotle (Fixed + ∃p) D1&2	(ii) Johnson-Laird (Free + ∃p) J-LS&B	(iii) Frege (Free − ∃p)
AA1	A(I)	A(I)	A
AA3	I	I	
AA4	I	A(I)	A
AE1		E	
EA1	E	E	E
AE2	E	E	E
EA2	E	E	E
AE3		O	
EA3	O	O	
AE4	E	E	E
EA4	O	O	
IA4	I	I	I
IA3	I	I	I
AI3	I	I	I
AI1	I	I	I
EI1	O	O	O
IE1		O	O
OA2		O	O
EI2	O	O	O
IE2		O	O
AO2	O	O	O
OA3	O	O	O
EI3	O	O	O
IE3		O	O
AO3		O	O
EI4	O	O	O
IE4		O	O

mative statements are the ones that surprise us the most if they turn out to be true. Our computational analysis below specifies the following ordering: A > M > F > I > E ≫ O (where ''>'' stands for ''more informative than'').

There are three generation heuristics:

G1. The *min*-heuristic: Choose the quantifier of the conclusion to be the same as the quantifier in the least informative premise (the *min-premise*).

We will show that the most informative conclusions that can validly follow from a pair of syllogistic premises almost always follows this rule. Furthermore, some conclusions probabilistically entail (''*p*-entail'') other conclusions. For example, if All X are Y, then it is probable that Some X are Y (this will follow as long as there are some Xs). We call these additional conclusions ''*p*-entailments.'' Thus, the second heuristic is:

G2. *P*-entailments: The next most preferred conclusion will be the *p*-entailment of the conclusion predicted by the *min*-heuristic (the ''*min*-conclusion'').

The *p*-entailments include a family of heuristics corresponding to the probabilistic relationships between the various quantified statements that make up the premises of a syllogism and which we describe later on. Heuristics (1) and (2) specify the quantifier of the conclusion. The third heuristic, the *attachment*-heuristic, specifies the order of end terms in the conclusion:

G3. *Attachment*-heuristic: If the *min*-premise has an end-term as its subject, use this as the subject of the conclusion. Otherwise, use the end-term of the *max*-premise as the subject of the conclusion.

We illustrate these heuristics with some examples. Consider AI1:

All Y are X	(*max*-premise)
Some Z are Y	(*min*-premise)
I-type conclusion	(by *min*)
Some Z are X	(by *attachment*)

By the *min*-heuristic, the conclusion is I. The *min*-premise has an end term (Z) as its subject. Therefore, by *attachment,* the conclusion will have Z as its subject term and the form *Some Z are X.* In contrast, consider AI4, where the order of terms in both premises is reversed and the *min*-heuristic also specifies an I conclusion. But now the I premise does not have an end term (neither X nor Z) as its subject. Therefore, the end term of the *max*-premise (i.e., All X are Y) is used as the subject in the conclusion, giving the form *Some X are Z.*

Now consider the following syllogism, IE2:

Some X are Y	(*max*-premise)
No Z are Y	(*min*-premise)
O-type conclusion	(by *p*-validity or logic, or by *p*-entailments)
Some X are not Z	(by *attachment*)
E-type conclusion	(by *min*)
No Z are X	(by *attachment*)

This is an IE syllogism where the *min*-heuristic does not give the type of the *p*-valid conclusion. The logical and *p*-valid conclusion is of type O. Here the conclusion order varies depending on whether the *min*-heuristic conclusion or the *p*-valid conclusion is chosen. Both logical and *p*-validity specify that the conclusion is of type O, whereas the *min*-heuristic specifies a conclusion of type E.

Logical validity and *p*-validity license an O conclusion, which contains the quantifier *Some.* The statements *Some* ϕ *are* ψ and *Some* ϕ *are not* ψ

have the same subject noun phrase (*Some* ϕ), which contains the same quantifier *Some*. More generally, the subject terms of O and I statements (whether as premises or conclusions) will attach or fail to attach in the same way. In this example, the quantifier in the conclusion is *Some*, which attaches to the quantifier of only one premise (the *max*-premise); moreover, the subject term of this premise is an end term (X), and so X is used as the subject term of the conclusion, leading to the conclusion *Some X are not-Z*. The *min*-heuristic licenses an E conclusion, which attaches directly to the second premise, producing the *No Z are X* conclusion. This makes the novel prediction that, in these cases, conclusion order will depend on conclusion type.

G1–G3 generate syllogistic conclusions. As noted above, we assume that people are unable to test these conclusions for *p*-validity (or, for that matter, logical validity). However, we assume that people employ two test heuristics that provide a fast and frugal estimate of how likely the conclusion generated by G1–G3 is to be informative and *p*-valid:

T1. The *max*-heuristic: Be confident in the conclusion generated by G1–G3 in proportion to the informativeness of the most informative premise (the *max*-premise).[3]

T2. The *O*-heuristic: Avoid producing or accepting O-conclusions, because they are so uninformative relative to other conclusions.

As we shall see in the computational level analysis, T1 gives a fast and frugal estimate of the likelihood that a syllogism has an informative, valid conclusion. T2 gives a fast and frugal heuristic for focusing on informative conclusions.

We now explain how we can justify these algorithmic level proposals at the computational level. In the Discussion we will show that these heuristics can be motivated intuitively and also that variants of them arise in other areas of reasoning.

Computational Level

The computational level analysis has three parts. Given the uncertainty of everyday reasoning, we first interpret quantified statements as probabilistic statements. This is consistent with our general probabilistic approach to other modes of reasoning, discussed above (Oaksford & Chater, 1998a, 1998b). Second, we use this probabilistic semantics in deriving an analysis of informativeness, which justifies the informativeness ordering over which the heuristics above are defined. We use Shannon's notion of ''surprisal'' (see (1) below) as our measure of information (Shannon & Weaver, 1949). The third part of the computational level analysis consists of an account of validity, which is defined over probabilistic statements. This probabilistic notion of validity (*p*-validity) has the advantage of allowing a definition of valid syllogistic inferences involving Most and Few, in a way that is uniform with a

[3] A slightly more exact formulation is given in the computational level analysis below.

definition of validity for the standard quantifiers.[4] Thus, we assume that people do not treat even syllogistic reasoning as a logical task, but rather that they assimilate it into their everyday probabilistic reasoning strategies. The notion of *p*-validity is crucial to generalizing the account to reasoning with Most and Few because, without a notion such as *p*-validity, we have no way of defining the correct answers to these generalized syllogisms. This explains why experimental studies of syllogistic reasoning with Most and Few have not previously been conducted, despite the fact that in, for example, mental models theory, there has been considerable theoretical interest in these quantifiers (e.g., Johnson-Laird, 1983, 1994).

Probabilistic semantics for the quantifiers. Quantified statements can be given intuitively natural meanings in terms of constraints on the conditional probability of the predicate (Y) given the subject (X).[5] ''All X are Y'' means that the probability of Y given X is 1 ($P(Y|X) = 1$). ''Most X are Y'' means that the probability of Y given X is high but less than 1 ($1 - \Delta \leq P(Y|X) < 1$, where Δ is small). ''Few X are Y'' means that the probability of Y given X is small but greater than 0 ($0 < P(Y|X) \leq \Delta$). ''Some X are Y'' means that the probability of Y given X is greater than 0 and that there are some things that are both Xs and Ys ($P(Y|X) > 0, \exists X, Y$). ''No X are Y'' means that the probability of Y given X is 0 ($P(Y|X) = 0$). ''Some X are not Y'' means that the probability of Y given X is less than 1 and that there some things that are Xs but not Ys ($P(Y|X) < 1, \exists X, \overline{Y}$).

This probabilistic account immediately provides the justification for the G2 heuristic. There are constraints between different probabilistic statements, which we show graphically in Fig. 2. This figure illustrates that there are inclusion relationships between the intervals associated with the different quantified statements. These inclusion relationships license the inferences that we call *p*-entailments in the G2 heuristic. A is included in I, i.e., $A \subset I$ which licenses the inference that if All X are Y, then Some X are Y (which holds just as long as there are some Xs), i.e., $A \Rightarrow I$. Similarly, $M \Rightarrow I$; $F \Rightarrow I$; $M \Rightarrow O$; $F \Rightarrow O$, and $E \Rightarrow O$. O and I overlap almost completely, aside from the endpoints, i.e., $O \Rightarrow I$ and $I \Rightarrow O$. Finally, there are weak relations, such that I and O are both compatible with M or F (although they

[4] M- and F-statements have been discussed in the literature on formal semantics (Barwise & Cooper, 1981), but no syntactic proof theory specifying what inferences follow from these statements has resulted from this work.

[5] This general approach was suggested to us by John R. Anderson (personal communication), and the possibility of a probabilistic approach to generalized quantifiers is mentioned in Moxey and Sanford (1991) and Johnson-Laird (1994). This approach has also been developed as a semantics for conditional statements, *if . . . then . . .* , by Adams (1966, 1975; but see Lewis, 1976, and Eells & Skyrms (1994) for discussion). Similar probabilistic semantics have been proposed in artificial intelligence by Pearl (1988). It is important to note, however, that we do not attempt here the much more difficult task of providing a full compositional semantics for quantified statements.

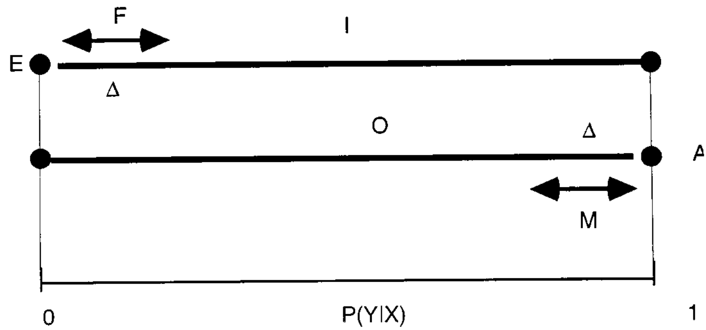

FIG. 2. The probabilistic semantics for the quantifers AMFIEO.

do not imply them). These relations depend on the size of Δ and the particular presumed values of P(Y|X), giving, I ⇒ m, f; and O ⇒ m, f (where lower case denotes weak *p*-entailments). Although some of the these relationships can be motivated on pragmatic grounds, all those involving M and F constitute novel predictions of our probabilistic semantics.[6]

We now apply this probabilistic semantics in deriving an informativeness ordering over quantified statements.

Informativeness. We now outline the notion of informativeness, which justifies the O-heuristic (T2) and underlies the other heuristics. We use information theory (Shannon & Weaver, 1949), where the informativeness $I(s)$ of a statement, *s*, is inversely related to its probability, $P(s)$:

$$I(\mathrm{s}) = \log_2 \left(\frac{1}{\mathrm{P(s)}} \right) \tag{1}$$

On our semantics for quantified statements, some relationships between informational strengths can be inferred directly, i.e., if $\Theta \Rightarrow \Phi$ then $I(\Phi) < I(\Theta)$,[7] providing a partial order corresponding to the *p*-entailments: $I(\mathrm{A}) > I(\mathrm{I})$; $I(\mathrm{M}) > I(\mathrm{I})$; $I(\mathrm{F}) > I(\mathrm{I})$; $I(\mathrm{M}) > I(\mathrm{O})$; $I(\mathrm{F}) > I(\mathrm{O})$; $I(\mathrm{E}) > I(\mathrm{O})$. But applying the *min*-heuristic requires a total order.

[6] Many researchers on syllogistic reasoning (e.g., Begg & Harris, 1982; Newstead, 1989; Newstead & Griggs, 1983; Politzer, 1986, 1990; Rips, 1994) assume that people interpret premises and conclusions not only in terms of logical form, but also in terms of conversational implicatures (Grice, 1975). We do not frame our discussion in terms of conversational pragmatics, although *p*-entailments may capture some effects usually discussed under this heading (e.g., Rips, 1994). These additional *p*-entailments are an important novel prediction of PHM.

[7] Existential presuppositions, in principle, complicate the account, but do not affect the derivation here because they are very uninformative. This is because almost any property Θ mentioned in discourse has some members. Even apparently nonreferring expressions, such as unicorns and fairies, may be referring in appropriate fictional domains.

Oaksford and Chater (1994) made a ''rarity'' assumption, i.e., the properties referred to in natural language typically apply to a small proportion of possible objects. Rarity affects the frequency with which different quantified statements will be true and, hence, their informativeness. Intuitively, rarity is important because most natural language predicates are mutually exclusive, e.g., *toupee* and *table* do not cross classify objects in the world. Therefore, the joint probability of being a toupee and a table is 0, i.e., the E-statement *No toupees are tables* is true. Because of rarity, this holds in general, i.e., an E-statement will typically describe the relationship between any two natural language properties. Therefore, E statements are almost always true and are, hence, uninformative. If most predicates applied to more than half the objects in the world, then their extensions would *have* to overlap and E-statements would be very surprising and informative. In Appendix A, we show how a general rarity assumption enforces the total order $I(\mathrm{A}) > I(\mathrm{M}) > I(\mathrm{F}) > I(\mathrm{I}) > I(\mathrm{E}) \gg I(\mathrm{O})$.

The O-heuristic is justified because O-statements are so uninformative. This is because, for almost all pairs of predicates X and Y, the statement Some X are not Y will be true (e.g., *Some toupees are not tables*). The analysis in Appendix A shows that O-statements are very much less informative than all other types of quantified statements.

Probabilistic validity. To justify the *min*-hueristic (G1), the *attachment*-heuristic (G3), and the *max*-heuristic (T1), we require a notion of validity which applies to probabilistic statements. We can then assess whether these heuristics favor *p*-valid conclusions.

Syllogistic reasoning relates two end terms *X* and *Z* via a common middle term *Y*. According to the probabilistic interpretation of the quantifiers, the two premises correspond to constraints on the conditional probabilities between the end terms and the middle term. We make the standard assumption (Pearl, 1988) of conditional independence between the end terms, given the middle term (i.e., we assume that the end terms are only related by the middle term). A *p*-valid conclusion follows if the premises place sufficient constraints on the conditional probability of one end term given the other (i.e., either $\mathrm{P}(Z|X)$ or $\mathrm{P}(X|Z)$). For example, if one of these probabilities is constrained to be 1, an A conclusion follows; if the conditional probabilities are constrained to be greater than 0, then an I conclusion follows, and so on. In Appendix B, we show how to derive the *p*-valid conclusions in detail. We have carried out these derivations for all 144 syllogisms involving the quantifiers AMFIEO. Table 2 shows the *p*-valid syllogisms involving the quantifiers AIEO, and Table 3 shows the additional *p*-valid syllogisms involving M and F.

We now compare our heuristics (G1, G3, T1) with *p*-validity to establish that they reliably predict the *p*-valid conclusion when there is one, thereby justifying the heuristics as a way to draw *p*-valid syllogistic conclusions.

First, we look at the predictions of the *min*-heuristic (G1) defined over

TABLE 2

The *p*-Valid Syllogisms Using the Standard AIEO Quantifiers, Showing the *p*-Valid Conclusion, the *p*-Valid Response Available in the Forced Choice Task (Dickstein, 1974), and the Conclusion Predicted by the *Min*-Heuristic

Syllogism	*P*-Valid conclusion	Valid conclusion available response	*Min*-heuristic
AA1	A	A	A
AA4	A		A
AI1	I	I	I
AI3	I	I	I
AE2	E	E	E
AE4	E	E	E
AO2	O	O	O
AO3	O		O
IA3	I	I	I
IA4	I	I	I
II1	I	I	I
II2	I	I	I
II3	I	I	I
II4	I	I	I
IE1	O		E
IE2	O		E
IE3	O		E
IE4	O		E
IO4	O		O
EA1	E	E	E
EA2	E	E	E
EI1	O		E
EI2	O		E
EI3	O		E
EI4	O		E
OA2	O		O
OA3	O	O	O
OI1	O	O	O
OO1	O	O	O
OO2	I	I	O
OO4	O		O

our standard information order. For AIEO, the type of the *p*-valid conclusion exactly fits the *min*-heuristic in 22 out of the 31 *p*-valid syllogisms (see Table 2). Eight of the nine violations involve I and E premises leading to an O conclusion, i.e., they lead to a *weaker* conclusion than predicted by the *min*-heuristic (O instead of E). Consequently, for 30 out of the 31 syllogisms, the *min*-heuristic provides an upper bound on the informativeness of conclusions. It seems unlikely that any cognitive inferential heuristics will be adapted to deriving O conclusions, because they are so uninformative. This leads to two predictions. First, people should be poorer at deriving valid O-

TABLE 3
The Additional *p*-Valid Syllogisms When the Standard AIEO Quantifiers Are Supplemented by M and F, Showing the *p*-Valid Conclusion, the *p*-Valid Response Available in the Forced Choice Task Used in Our Experiments, and the Conclusion Predicted by the *min*-Heuristic

Syllogism	Valid conclusion	Valid conclusion available response	*Min*-heuristic
AM1	M ∨ A	M ∨ A	M
MA1	M	M	M
MM1	M	M	M
AM2	O	O	M
MA2	O		M
MM4	M		M
AM4	M		M
MA4	M ∨ A		M
FA1	F	F	F
FM1	F	F	F
AF2	F ∨ E	F ∨ E	F
FA2	F ∨ E		F
MF4	F		F
AF4	F		F
MF1	O	O[AMFO]	F
MO1	O	O	O
FF1	O	O[AMFO]	F
FO1	O	O	O
OM1	O	O	O
OM3	O	O	O
OF1	O	O	O
OF3	O	O	O
FO3	O		O
FO4	O		O
MO3	O		O
MO4	O		O
OF4	O		O
FF4	O		F
OM4	O		O
FM4	O		F
MI1	I, O	I	I
MI3	I, O	I	I
IM3	I, O	I	I
IM4	I, O	I	I
FI1	I, O	I	I
FI3	I, O	I	I
IF3	I, O	I	I
IF4	I, O	I	I

Note. The notation ‘‘M ∨ A’’ indicates the constraint on the relevant conditional probability is that it must be greater than $1 - \Delta$ and, consequently, that either M or A follows. Similarly, ‘‘F ∨ E’’ indicates that the relevant conditional probability must be less than Δ and, consequently, that either F or E follows. The notation I, O indicates that the conditional probability must be greater than 0 and less than 1 and, therefore, that both conclusions can follow. The superscript AMFO indicates that this syllogism occurs in both our subsequent experiments but that this response option is only available in the experiment using just the AMFO quantifiers (see the section on Experiments on Generalized Quantifiers).

conclusions than other valid conclusion types. Second, within the syllogisms with valid O-conclusions, people should find those conforming to the *min*-heuristic the easiest. The one remaining exception to the *min*-heuristic is OO2, for which the conclusion I is *p*-valid, whereas the *min*-heuristic recommends the less informative O-conclusion. The *p*-validity of this conclusion depends essentially on the independence assumption—because logical validity does not involve this assumption, this inference is not logically valid. Thus, our account of *p*-validity establishes that the *min*-heuristic is, indeed, a reliable method of determining the form of the *p*-valid conclusion.

Of the 80 additional syllogisms obtained by introducing M and F, 38 are *p*-valid. For these, the type of the *p*-valid conclusion exactly fits the *min*-heuristic in 32 cases (Table 2). All six violations (MF1, FF1, AM2, MA2, FM4, FF4) have O conclusions, which are *weaker* conclusions than the *min*-heuristic predicts. Consequently, for all 38 *p*-valid syllogisms the *min*-heuristic provides an upper bound on conclusion informativeness. Overall, 54 of the 69 *p*-valid syllogisms conform exactly to the *min*-heuristic, 14 have less informative *p*-valid conclusions, and only one violates the *min*-heuristic. Thus, *p*-validity confirms that the *min*-heuristic is reliable.

Second, we show that the *attachment*-heuristic (G3) derives the correct conclusion orders for almost all syllogisms and so can be justified as a method of picking out *p*-valid conclusions. The only relevant syllogisms are those with asymmetric conclusions, where conclusion order matters. Attachment correctly predicts all 14 of the AIEO syllogisms with asymmetric logically valid conclusions and 44 of the 48 AMFIEO syllogisms with asymmetric *p*-valid conclusions (the exceptions are MF1, FM4, AM2, and MA2). Therefore, *attachment* reliably gives the logically or *p*-valid conclusion order, where one exists.

There are a small number of syllogisms where *attachment* does not apply. Where the conclusion type either attaches to both or neither premise (i.e., for syllogistic figures 2 and 3), *attachment* decides the conclusion order from the most informative premise. But this criterion does not apply if the premises are of the *same* type (e.g., AA, MM, etc.) and so neither conclusion is preferred. With respect to the *min* conclusion, this arises for 12 syllogisms, of which 8 have asymmetrical conclusions (AA2, AA3, MM2, MM3, FF2, FF3, OO2, OO3). With respect to conclusions recommended by *p*-entailment, this arises for all 24 syllogisms where the premises are of the same type, irrespective of figure, leading to 12 asymmetric cases (MM1, MM2, MM3, MM4, FF1, FF2, FF3, FF4, EE1, EE4, OO1, OO4). Although *attachment* does not apply to all syllogisms, it makes predictions for figure 2 and 3 syllogisms made by no other theory. Moreover, it provides a reliable way of determining conclusion order for most syllogisms.

Finally, we justify the *max*-heuristic which provides a fast and frugal test-heuristic for assessing confidence in the *p*-validity of the candidate conclusion generated by the *min*-heuristic. The *max*-heuristic is based on the *most* informative (i.e., the *max*) premise and suggests that confidence is proportional to the average informativeness of the conclusions of syllogisms with that *max*-premise. For example, consider AO and EO, which both generate an O-conclusion. However, the *max* premise type for AO is A which has high average informativeness, while for EO it is E which has low average informativeness. Thus, confidence that O follows validly from AO should be higher than from EO and so, according to the *max*-heuristic, O should be selected in far higher proportion for AO than for EO syllogisms.

We now show how to compute average informativeness, thereby justifying the *max*-heuristic. Of the 144 AMFIEO syllogisms, 44 syllogisms have *max* premise type A, 36 have M, 28 have F, 20 have I, 12 have E, and four have O. We counted the number of conclusions recommended by the *min*-heuristic where there is a possible informative conclusion (i.e., those that are also *p*-valid). For example, for the 44 syllogisms with *max* premise type A, the *min*-heuristic recommends the A conclusion in four cases, of which two are *p*-valid. All other conclusion types are recommended by the *min*-heuristic eight times, of which the following numbers are *p*-valid: M = 6, F = 4, I = 4, E = 4, and O = 6. To calculate the expected informativeness, we weight each *min*-conclusion (and its *p*-entailments) which is *p*-valid by the informativeness of that conclusion type (and its *p*-entailments), which can be estimated from our model in Appendix A ($I(A) = 4.99 > I(M) = 4.35 > I(F) = 3.00 > I(I) = 1.83 > I(E) = .48 > I(O) = .05$). *Min*-conclusions that are not *p*-valid are assigned 0 informativeness, because no information is carried by conclusions which do not follow from the premises. The expected informativeness, *ExpInf,* of the syllogisms with the different *max* premise types are: *ExpInf*(A) = 1.99; *ExpInf*(M) = 1.41; *ExpInf*(F) = 1.02; *ExpInf*(I) = .76; *ExpInf*(O) = .05; and *ExpInf*(E) = 0. This predicts that confidence in the *min*-conclusion should follow the following order in *max*-premise types: $A > M > F > I > O \approx E$.

Although it may be possible to justify other more complex heuristics for confidence, we have chosen to explore a simple heuristic, which concentrates on a single reliable cue rather than integrating multiple cues, in potentially complex ways. Gigerenzer and Goldstein (1996) suggested that the cognitive system often uses nonintegrative but ‘‘fast and frugal’’ heuristics of this kind (see also Gigerenzer, Hoffrage, & Kleinbölting, 1991).

Summary and Preview

The first part of this paper outlined and justified PHM. We have shown that the *min*- and *attachment*-heuristics almost always yield a correct informative conclusion, *if* there is one. Because these strategies can be applied to all syllogisms, whether there is a *p*-valid conclusion or not, systematic errors will arise for those syllogisms that do not have a valid conclusion. Test procedures, which lie outside the scope of PHM, may be able to distinguish valid from invalid syllogisms to some degree. However, these are complex processes likely to be subject to large individual variation. The *max*-heuristic provides a ''fast and frugal'' (Gigerenzer & Goldstein, 1996) estimate of the likely validity and potential informativeness of a syllogistic conclusion. In conclusion, according to PHM, although people are poor at distinguishing valid from invalid syllogisms, they can get the valid conclusion if there is one. We shall find below that overgeneralization according to the *min*-heuristic is extremely widespread in laboratory tasks.

PHM follows recent work showing that simple heuristics can be highly adaptive insofar as they approximate optimal solutions (Gigerenzer & Goldstein, 1996; McKenzie, 1994). These heuristics are used in place of complex and sometimes computationally intractable optimal strategies, which are therefore denied any cognitive role. The role of the optimal solution is to explain why the heuristic is adaptive. Our theory of syllogisms follows this pattern of reasoning. The principal difference is simply that we develop the optimal theory and processing heuristics in tandem, rather than seeking a post hoc explanation of the success of processing principles observed in human performance.

In the next three sections of the paper, we evaluate the empirical adequacy of PHM. We first compare the predictions of PHM with the results of a meta-analysis of existing data on syllogistic reasoning. We then compare the adequacy of PHM with a range of alternative theoretical accounts. Finally, we describe two experimental tests of the novel predictions of PHM with syllogisms using *Most* and *Few*.

META-ANALYSIS

We conducted a meta-analysis using the five experiments that used all 64 syllogisms—Dickstein's (1978) first test (D1, $N = 22$) and second test (D2, $N = 76$); Johnson-Laird and Steedman's (1978) first test (J-LS1, $N = 20$) and second test (J-LS2, $N = 20$); and Johnson-Laird and Bara's (1984) experiment 3 (J-LB, $N = 20$). The response format varied in these studies. In Dickstein (1978), participants made a forced choice between A, I, E, O and ''no valid conclusion'' (NVC) responses. The conclusion was always presented in its traditional Z–X order, reducing the number of valid syllogisms from 27 to 19 (see Table 1). The Johnson-Laird studies used a ''production task,'' where participants are asked to ''write down what followed from each pair of statements.'' In testing many of the predictions of PHM, we are only concerned with the form of the conclusion, i.e., is it A, I, E, O, or NVC? Therefore, we recast all the Johnson-Laird data into these five response categories disregarding the X–Z or Z–X conclusion order.[8] In analyzing these

[8] This meant that on asymmetric conclusion types (A and O), some logically invalid responses were classified as correct. However in J-LS1, J-LS2, and J-LB, this applied to only 7.9% of participants' responses for asymmetric conclusions.

results, we use the concept of logical validity used in modern mathematical logic, where syllogisms requiring an existential presupposition are not valid (Table 1(iii)).

The procedural change between the forced choice and production tasks could affect our analysis. There are eight syllogisms (see Table 1) that have a valid conclusion which is not presented in the forced choice task. However, for these eight syllogisms, there are high product moment correlations for the frequencies with which participants draw particular conclusion types: the correlations ranged between $r(38) = .43$ and $r(38) = .92$ (mean = .72). Importantly, the lowest correlation was *not* between the Dickstein studies and the Johnson-Laird studies, but between J-LS2 and J-LB which both used the *same* response mode. Consequently, the change of response mode appears to have had little effect on the form of the conclusion that participants draw and, hence, it is reasonable to combine these studies in our meta-analysis.

In this section, we look first at the overall fit between PHM and the data. We then turn to more fine-grained analyses testing the detailed predictions of PHM.

Modeling

Appendix C shows the means (weighted by sample size) of each conclusion type for each syllogism. Responses predicted by the *min*-heuristic are underlined and in italics. Those that do *not* rely on *p*-entailments are also in bold. Appendix C is qualitatively very revealing. First, the *min*-heuristic almost invariably predicts the two modal responses for each syllogism type (with an apparent violation on EO, which we discuss below). PHM accounts for 96.75% of responses (excluding NVC). Second, the response format in the forced-choice and production tasks precludes more than one response. Consequently, participants are unlikely to make *min*-responses that rely on *p*-entailments. This predicts that: (i) there should be more *min*-responses that don't rely on *p*-entailments than do rely on *p*-entailments, and (ii) there should be more *min*-responses that rely on *p*-entailments than responses not predicted by the *min*-heuristic. Appendix C bears out both predictions. Third, in Appendix C, there is a clear linear order in the frequency of responding by *max*-premise type such that A $>$ I $>$ E $\approx$ O as predicted by the *max*-heuristic. Moreover, Appendix C shows that the NVC response is inversely proportional to confidence, as the *max*-heuristic predicts. In sum, Appendix C reveals that PHM seems to capture the majority of the effects in the data.

We now assess the overall fit between data and model by adding some parameters to PHM which attach specific numerical values, rather than a simple linear order, to the degrees of confidence suggested by the *max*-heuristic. Each parameter indicates the percentage of choices according to the *min*-heuristic for each *max*-premise type: $p_A(70.14)$, $p_I(31.11)$, $p_E(18.78)$, $p_O(18.04)$. We assume that participants draw *p*-entailments on a fixed proportion of syllogisms, which is the same for all types. $p_{ent}(10.76)$ indicates the percentage of *p*-entailment responses averaged over all syllogism types. Finally, $p_{error}(1.22)$ indicates the percentage of responses not predicted by the *min*-heuristic including *p*-entailments. We obtained best fit estimates of these six parameters directly from the mean values of the relevant quantities in the data shown in Appendix C (see above the values in parentheses).

We computed the fit between data and model over all five responses, A, I, E, O, and NVC, for all 64 syllogisms, making 320 data points in all (see Appendix C). Estimates for the NVC response did not involve introducing any further parameters because they can be derived from a linear combination of the existing parameters. The product moment correlation between the 320 observed and predicted values was good, $r(318) = .90$, and the root-mean-square deviation was .30 (logit–logit fits were assessed to control for the undue influence of extreme data points; see Belsey, Kuh, & Welsch, 1980). Thus, PHM accounts for over 80% of the variance and so provides good overall fits to the empirical data on syllogistic reasoning. In the following sections, we turn to more fine-grained meta-analyses of the predictions of PHM.

The Min-Heuristic and Conclusion Informativeness

We now consider the extent to which PHM's success may be parasitic on capturing logical validity. To do this, we consider the valid and invalid syllogisms separately.

Logically valid syllogisms. According to our informational analysis, O conclusions are much less informative than A, I, and E conclusions. Consequently, for the valid syllogisms, there should be more correct solutions for the A, I, and E than for the O conclusions. Appendix C shows the frequency with which each conclusion type was drawn for all syllogisms. The 22 valid syllogisms are shown in italics. Appendix C reveals that *all* the A, I, and E conclusions were drawn more frequently than *all* the O conclusions. This difference was reliable both by study, $t(4) = 6.11$, $p < .005$ (means: AIE = 86.33 (*SD:* 4.41); O = 47.07 (*SD:* 18.17)), and by materials, $t(20) = 7.34$, $p < .0001$ (means: AIE = 86.33 (*SD:* 4.47); O = 47.07 (*SD:* 16.36)).[9] For the A, E, and I conclusions, the *min*-heuristic works perfectly. For the O conclusions, the *min*-heuristic works for the AO (and OA) syllogisms. However, it only captures the O response for the EI (or IE) syllogisms via *p*-entailments. The *min*-heuristic, therefore, predicts that participants should

[9] Throughout this paper, analyses by study are computed as within subject analyses, whereas analyses by materials are computed as between subject analyses because the *N*s are invariably unequal by materials. Furthermore, for these studies, the means are unweighted by sample size for both analyses. In general, throughout our meta-analyses, where possible, we have weighted means by sample size (Wolf, 1986). Occasionally, this means that there are discrepancies between the same means reported in different analyses.

find the AO syllogisms easier than the EI syllogisms and Appendix C reveals that, except for EI1, the valid O conclusion was drawn more frequently for *all* the AO syllogisms than for *all* the EI syllogisms. This difference was reliable both by study, $t(4) = 4.76$, $p < .01$ (means: AO = 61.26 (*SD:* 20.83); EI = 39.97 (*SD:* 17.66)), and by materials, $t(10) = 2.64$, $p < .025$ (means: AO = 61.26 (*SD:* 6.23); EI = 39.97 (*SD:* 15.21)).

Logically invalid syllogisms. We first show that the *min*-heuristic predicts participants' responses when they incorrectly assume that there is a valid conclusion. Excluding the valid syllogisms, the *min*-heuristic still accounts for 95.17% of conclusions drawn (excluding NVC). If people are overgeneralizing to the invalid syllogisms, then retaining the parameter values we derived from the overall data should provide as good a fit to the data on invalid syllogisms alone. $r(208) = .90$ (rms error = .29) which is the same level of fit achieved for the overall data. Consequently, the same informational strategy accounts for the valid and invalid syllogisms.

The Max-Heuristic

Min responses. The *max*-heuristic predicts that there should be a linear order in the frequency of the *min*-heuristic response dependent on the *max*-premise such that A > I > O ≈ E (see Appendix C). A one-way ANOVA with *max*-premise type (A, I, E, O) as the within studies factor and the percentage of responses that the *min*-heuristic predicts as the dependent variable revealed a highly significant linear contrast in the predicted direction, $F(1, 12) = 120.39$, $MS_e = 59.62$, $p < .0001$, confirming that confidence in the *min*-response is determined by the *max*-heuristic, as PHM predicts.

No valid conclusion responses. PHM predicts that the less confident participants are in the *min*-heuristic conclusion the more NVC responses they should make. Consequently, confidence and the frequency of NVC responses should be inversely related. We tested this prediction using a one-way ANOVA with *max*-premise type (A, I, E, O) as the within studies factor and the percentage of NVC responses as the dependent variable (see Appendix C). There was a highly significant linear contrast in the predicted direction, $F(1, 12) = 173.85$, $MS_e = 49.46$, $p < .0001$, confirming that participants' confidence in the *min*-heuristic conclusion is determined by the *max*-heuristic and that they are more likely to make the NVC response when confidence is low, as PHM predicts.

P-*Entailments*

PHM predicts that (i) there should be more *min*-heuristic responses that do not rely on *p*-entailments (MIN) than do rely on *p*-entailments (MIN′), and (ii) there should be more MIN′ responses than responses that the *min*-heuristic does not predict. To test prediction (i) within each study for each syllogism, if the number of participants making the MIN response was greater than the number of participants making the MIN′ response, then we classified a syllogism as *min*. Conversely, if the number of participants making a MIN′ response was greater than the number of participants making the MIN response, then we classified a syllogism as *ent*. (We excluded a syllogism when there was a tie for the number of participants endorsing each response.) As PHM predicts, there were more *min* syllogisms (across the five studies, the range was 50 to 56) than *ent* syllogisms (across the five studies, the range was 5 to 13). We performed a directly analogous analysis to test prediction (ii). For each syllogism, if the number of participants making the MIN′ response was greater than the sum of the number of participants selecting either of the responses that the *min*-heuristic does *not* predict, then we classified the syllogism as *ent*. Conversely, if the sum of the number of participants selecting either of the responses that the *min*-heuristic does *not* predict was greater than the number of participants making the MIN′ response, then we classified the syllogism as *non-min*.[10] As PHM predicts, there were more *ent* syllogisms (across the five studies, the range was 22 to 36) than *non-min* syllogisms (across the five studies, the range was 0 to 16). For both predictions (i) and (ii), all differences were highly significant for each study.

The Attachment-Heuristic

In this section, we test the predictions of the *attachment*-heuristic. This heuristic can only be assessed for the data from J-LS1, J-LS2, and J-LB where both conclusion orders are possible. The *attachment*-heuristic predicts the conclusion order both of the *min*-conclusion and of its *p*-entailments. For the *min*-conclusions, we considered each syllogism in turn and compared the number of participants producing that conclusion in the order predicted by attachment with the number of participants producing the opposite conclusion order. There are 56 conclusions for which attachment makes a prediction. Of these, 54 syllogisms conformed to attachment, 1 violated attachment, and there was 1 tie. There are 52 conclusions for which attachment makes a prediction for the *p*-entailments. Of these, 34 syllogisms conformed to attachment, 1 violated attachment, and there were 17 ties. Removing the logically valid syllogisms did not affect this result. Interestingly, the result was also unchanged when focusing on symmetric conclusions, i.e., where the *min*-response is E or I. Attachment still predicts a systematic bias towards a particular conclusion order for all figures. These results confirm that conclusion order is determined by the conclusion type, as the *attachment*-heuristic suggests.

The *min*- and *attachment*-heuristics also explains a well-known descriptive generalization about conclusion order: the ‘‘figural effect.’’ This is the tendency to select the end term that is the subject of one of the premises (if

[10] With the exception of EO syllogisms, where we compared the mean of the two responses predicted by MIN_{imp} with the response not predicted by the *min*-heuristic.

there is exactly one) as the subject of the conclusion. For example, for Fig. 1, Y–X, Z–Y, the end term Z is the only end term in subject position so, by the figural effect, Z should be selected as the subject of the conclusion, which therefore has the order Z–X. Conversely, in syllogistic figure 4, X–Y, Y–Z, the end term X is the only end term in subject position so the conclusion order is X–Z. The figural effect cannot apply to syllogistic figures 2 and 3 because in neither figure is just one end term in subject position in a premise (its either neither or both end terms). The figural effect is purely descriptive—it has not been proposed to have any rational basis. However, mental models theory (Johnson-Laird, 1983; Johnson-Laird & Bara, 1984) does have an explanation of the figural effect in terms of the processing characteristics of working memory, which we consider below.

For all 32 syllogisms in Figs. 1 and 4, the predictions of the *min*- and *attachment*-heuristics are exactly in agreement with the figural effect. Moreover, of the 24 syllogisms in Figs. 2 and 3 for which the *attachment*-heuristic makes predictions, 22 conformed to attachment, 1 violated attachment, and there was 1 tie. Removing the logically valid syllogisms again does not alter this result. In sum, in accounting for the data on conclusion order, the *attachment*-heuristic goes beyond a combination of the figural effect and logic.

A strong prediction of the *attachment*-heuristic, which cannot be captured by the figural effect, is that conclusion type determines conclusion order. Therefore, for some syllogisms, the conclusion order of the *min*-conclusion is predicted to be the opposite to that of its *p*-entailment. This arises for eight syllogisms: AE1, AE2, EA2, EA4, IE1, IE2, EI2, EI4. According to attachment, there should be an interaction between the choice of conclusion order between participants choosing the *min*-conclusion and those choosing the *p*-entailment. To assess this, within studies we used Binomial tests for simple effects and Fisher's exact tests for interactions, using syllogism as the unit of analysis. We then combined these tests meta-analytically using Stouffer's method (Wolf, 1986). Across studies, when the *min*-conclusion was drawn, more syllogisms conformed to the *min*-conclusion order than to the opposite order, $z = 1.73$, $p < .05$ (one-tailed); however, when the *p*-entailment was drawn, more syllogisms conformed to the *p*-entailment conclusion order than to the opposite order, $z = 2.74$, $p < .005$ (one-tailed). The interaction, i.e., the difference between these simple effects, was highly significant, $z = 2.69$, $p < .005$ (one-tailed). This analysis confirms that conclusion type influences conclusion order within the same syllogism, which is predicted only by the *attachment*-heuristic. It also confirms that conclusion order is decided after conclusion type is selected.

Modeling. We now model the data including conclusion order in the same way as we modeled conclusion type. As before, we require four parameters, $p_A(55.30)$, $p_I(24.08)$, $p_E(15.97)$, $p_O(15.00)$, indicating the percentage of choices according to the *min*-heuristic for each *max*-premise type (A, I, E, O). These parameters are calculated with respect to the conclusion order predicted by attachment. p_{ent}(8.62) now indicates the percentage of choices involving *p*-entailments averaged over all syllogism types, but only for the correct conclusion order. For symmetrical conclusion types (I and E) where order does not have any logical significance, we assume that participants will sometimes (although quite infrequently) select the conclusion order opposite to that predicted by attachment. The parameter p_{opp}(6.38) stands for the proportion of responses of this kind, whether the responses are predicted by the *min*-heuristic or by the *p*-entailments. p_{error} (.64) again indicates the percentage of errorful responses. We obtained best fit estimates of these 7 parameters directly from the mean values (see values in parentheses above). For the 8 syllogisms where the *attachment*-heuristic does not specify a conclusion order either for the *min*-conclusion or its *p*-entailment or both, we add together the two possible conclusion orders in estimating the best fit parameter. Having obtained this best fit value, we then apply it by assigning half the parameter value to each possible conclusion order. The product moment correlation between model and data was $r(510) = .86$ (rms error = 7.56). Thus, the model accounts for 74% of the variance in the data. This is a good fit given that just 7 parameters have been used to explain 512 data points.

In summary, the *attachment*-heuristic seems to account accurately for the data on conclusion order. Previous accounts rely on the figure of the premises to explain the observed biases. These accounts only work for figures 1 and 4. One of the virtues of mental models was that it could explain the effects for these figures by the presumed ''first-in-first-out'' characteristics of a putative working memory store (Johnson-Laird, 1983; Johnson-Laird & Bara, 1984). The *attachment*-heuristic not only explains the data on figures 1 and 4 but also the complex pattern of data from figures 2 and 3. Moreover, the *attachment*-heuristic correctly predicts the interaction between the conclusion type chosen and the conclusion order selected.

Logical Validity

Does logical validity play any residual role in syllogistic reasoning performance? In our meta-analysis, participants gave the logically correct answer on only 51.8% of these trials (''correct'' in terms of conclusion type *and* conclusion order for valid syllogisms and NVC for the invalid syllogisms[11]). This is considerably better than chance responding (in the forced choice task, chance = 20%; in the production task, chance = 11.1%). However, for the valid syllogisms, participants gave the logically correct answer on 64.9% of

[11] The Dickstein studies force participants to choose from the full range of conclusion types AIEO, but presented in only its classical order, Z–X, e.g., *All Z are X,* but not *All X are Z.* Therefore, for syllogisms that do have valid conclusions, but only in the X–Z order, the correct response is defined as NVC. This means that the Dickstein studies have only 16 valid syllogisms, as opposed to 22 in the other three studies.

trials, whereas for the invalid syllogisms, the figure was only 46.8%. Clearly, performance is better on the valid syllogisms. This pattern is consistent with PHM, which recommends the correct answer when there is one, but which also recommends an answer when there is no valid conclusion (although the *max*-heuristic modifies confidence in this conclusion). So to what extent are people sensitive to logical validity, over and above these heuristics, i.e., to what extent do logical test processes affect people's responses?

We looked first at differences in the frequency of responding for valid and invalid syllogisms controlling for the effects of the *min*- and *max*-heuristics. The A *max*-premise type syllogisms have equal numbers of valid and invalid syllogisms for each syllogism type (AA, AI, AE, AO). We tested for differences using the percentage of *min*-response as the dependent variable. For valid A *max*-premise type syllogisms, the *min*-response is always valid; but following the *min*-heuristic would equally give conclusions for these syllogisms that are invalid. Consequently, if participants are attending to logical validity, there should be more *min*-responses for the logically valid than for the logically invalid A *max*-premise type syllogisms. There were significant differences both by study, $t(4) = 3.16$, $p < .05$ (means: Valid = 79.16 (*SD:* 9.07); Invalid = 59.67 (*SD:* 8.39)), and by materials, $t(26) = 4.52$, $p < .001$ (means: Valid = 79.16 (*SD:* 14.33); Invalid = 59.67 (*SD:* 12.04)). Participants are generally more confident of their response when the conclusion is also logically valid. Consequently, it seems that even when the *min*- and *max*-heuristics are controlled for, our meta-analysis still reveals a logical validity effect (Revlis, 1975a, 1975b). Although this result is statistically significant, its psychological significance is less clear.

We therefore conducted more specific analyses on two critical sets of syllogisms where PHM and logic diverge on the predicted conclusion type. First, for the eight IE syllogisms the *min*-heuristic predicts an E-conclusion and logical validity predicts an O-conclusion. The *min*-heuristic also predicts people should draw an O conclusion but only if they draw the *p*-entailment. We assessed these contrary predictions using Fisher's exact tests (Siegel & Castellan, 1988) within studies and combining the results using Fisher's combined test (Wolf, 1986). Across studies, there was no significant difference in the number of IE syllogisms where O was preferred to E, than syllogisms where E was preferred to O, $\chi^2(10) = 14.26$, $p > .10$. This result suggests that logical test procedures only have a small influence on the output of the generate procedures. Nevertheless, the modal responses to IE syllogisms is O, which even taking *p*-entailments into account is not predicted by PHM.

In conclusion, putative logical test procedures do appear to have some influence alongside generate procedures in determining people's responses.

Nonpredicted Aspects of the Data

There are two aspects of the data not predicted by PHM. First, there are 20 syllogisms for which the *attachment*-heuristic does not predict a preferred conclusion order. However, there is a systematic pattern in participants' responses. Clearly, because the conclusion must be produced in one order or the other, participants must ''break symmetry,'' even when attachment offers no advice. Specifically, it appears that for these cases, people typically prefer the A–C order. Of the 20 possible cases, there are 6 ties, simply because no responses of this conclusion type are ever made. Of the remaining 14 cases, 12 have the A–C order as the predominant response, and only 2 have the C–A order as the predominant response. Although this effect may have no rational basis, it may reflect a simple processing strategy, which is used as a last resort. For example, the first-in-first-out principle (Johnson-Laird and Bara, 1984; Johnson-Laird & Byrne, 1991) could explain this effect. In a first-in-first-out memory, because the A term is encountered first and the C term second, when these items are retrieved to form a conclusion, the most natural retrieval order is A–C.[12]

Second, for the eight EO syllogisms, the E conclusion is sometimes selected, which is not predicted either by the *min*-heuristic or by its *p*-entailment. One possible explanation is that *p*-entailments can be made not only from conclusions, but also from the premises (see Rips, 1994). That is, perhaps *p*-entailments are sometimes drawn from one or the other or both of the premises and the *min*-heuristic then applied to the resulting premises. For the eight EO syllogisms, this would produce the E conclusion. If the *p*-entailment of the O premise is taken, this yields an I premise, and therefore an IE syllogism is generated. By the *min*-heuristic, the conclusion of this syllogism is of type E. Interestingly, for all the other nine syllogism types (i.e., the remaining 56 syllogisms), introducing *p*-entailments of the premises in this way does *not* produce any additional conclusion types. However, because adding *p*-entailments of the premises would complicate PHM, we do not include this possibility.

Summary

We have shown that the empirical data is consistent with PHM. The model shows good fits with *all* the data included in our meta-analysis. PHM also makes fine-grained predictions concerning the valid syllogisms, the invalid syllogisms, the *max*-heuristic, the NVC response, and the *p*-entailments. We confirmed all these predictions in our meta-analysis.

COMPARISON WITH OTHER THEORIES

We now compare PHM with leading alternative theories of syllogistic reasoning: Mental logic (Rips, 1994), mental models (Johnson-Laird, 1983;

[12] This is a more straightforward application of the first-in-first-out principle than used by Johnson-Laird and Bara (1984) because for them the order of input is not determined by the order of the premises, but by the way in which the participant decides to tackle the problem.

Johnson-Laird & Byrne, 1991), the deduction as verbal reasoning framework (Polk & Newell, 1988, 1995), and the atmosphere (Woodworth & Sells, 1935), matching (Wetherick, 1989), and conversion hypotheses (Chapman & Chapman, 1959).

Mental Logics: The Evaluation Task

The mental logic view is that people reason logically using formal rules like those of a natural deduction system (Braine, 1978; Galotti, Baron, & Sabini, 1986; Rips, 1983). Rips (1994) provides such an account of syllogistic reasoning, PSYCOP, and models his own data from a new task variant, where participants *evaluated* syllogisms, given both premises *and* the conclusion. Rips used a fixed order of terms in the conclusion, leading to four conclusion types for each syllogism, giving $4 \times 64 = 256$ problems in total. Participants indicated whether the conclusion was ''necessarily true'' or ''not necessarily true'' given the truth of the premises, i.e., whether the conclusion did or not did follow. The crucial novelty of the evaluation task is that it allows participants to endorse more than one possible conclusion, given the same premises.

Rip's PSYCOP is a computational model based on natural deduction. Rips allows for errors in two ways. First, participants may sometimes make ''conversational implicatures'' which go beyond the logical form of the premises (Begg & Harris, 1982; Newstead, 1989; Newstead & Griggs, 1983). These implicatures correspond closely to the *p*-entailments that follow from the probabilistic semantics for the quantifiers in PHM. Second, when participants cannot derive the conclusion, even taking conversational implicatures into account, they sometimes simply guess. Rips relates logical inference, conversational implicatures, and guessing into a complex flow diagram, where paths are associated with the probability that a subject uses that path. This introduces 10 independent parameters, which Rips estimates from his data. PSYCOP provides an excellent fit to Rips's data. We now show that PHM provides as good a fit.

Modeling the evaluation task. The main difference between Rips' data (summarized by conclusion type in Table 4) and the standard results (see Appendix C) is that the responses the *min*-heuristic predicts *with* *p*-entailments are now at similar levels to the responses predicted by the *min*-heuristic *without* *p*-entailments. This is not surprising because, in Rips' procedure, the response categories, A, I, E, O, are not mutually exclusive, so participants can make both MIN and MIN′ responses. This also means that we can perform a factorial analysis of the *p*-entailments. We performed a one-way ANOVA by materials, with response category as a within participants factor with three levels: MIN responses, MIN′ responses, and other responses. The dependent variable was the percentage of participants endorsing that response category. The means and standard deviations were as follows: MIN: 29.84 (24.43) MIN′: 25.31 (18.85), and not predicted by the

TABLE 4
Summary of Rips' (1994) Data Indicating the Mean Percentage of Each Conclusion Type Endorsed

	Conclusion-type			
Syllogism	A	I	E	O
AA	***46.25***	*46.25*	2.50	5.00
AI	2.50	***53.13***	1.88	*25.63*
AE	1.25	0.63	***54.38***	*42.50*
AO	1.25	*25.00*	2.50	***28.75***
II	1.25	***26.25***	0.00	*27.50*
IE	0.00	8.75	***6.25***	*38.13*
IO	0.63	*18.13*	0.63	***18.13***
EE	2.50	5.00	***21.25***	*8.75*
EO	1.25	*10.00*	*3.75*	***20.63***
OO	0.00	*20.00*	0.00	***21.25***

Note. The numbers that are in italics are those predicted by the *min*-heuristic with *p*-entailments, as outlined in the text. The numbers which are in bold are those which would also be predicted without *p*-entailments, as in the orginal analysis. The AI syllogism includes the IA syllogisms, the AE syllogisms include the EA syllogisms, and so on.

min-heuristic (non-min): 2.46 (2.87).[13] We used planned contrasts to test whether, as with the production and forced choice tasks, the order in response preference was MIN > MIN′ > other. There was no significant difference between MIN and MIN′ ($F(1, 126) = 2.44$, $MS_e = 269.81$, $p = .12$). We therefore compared these levels collapsed with non-min. There was a highly significant difference ($F(1, 126) = 99.76$, $MS_e = 269.81$, $p < .0001$). Thus, as in our meta-analysis, participants' main pattern of responses follow the *min*-heuristic. However, there is now no difference in the frequency of MIN and MIN′ responses, due to the change of procedure.

We now model Rips's data. This exactly mirrors our approach in our meta-analysis, except that the lack of a significant difference between the frequency of MIN and MIN′ means that we can drop p_{ent}, the *p*-entailments parameter. Therefore, we simply use the *max*-premise type parameters to model both responses. We estimate these parameters by averaging over all these values for each *max*-premise type in Table 4 ($p_A = 39.38$; $p_I = 21.50$; $p_E = 12.34$; $p_O = 20.63$; $p_{error} = 1.96$.). Using logit–logit fits to minimize the effects of extreme values as before, we obtained a very good fit to the overall data, $r(254) = .81$, rms error $= .44$.[14] For comparison, we computed

[13] This last mean is not the same as p_{error}, because it is computed over all the data and not over the summary data.

[14] Rips data did not conform to the A > I > E > O confidence order by *max*-premise type. To confirm the fit with this order imposed, we derived estimates by performing a regression analysis between the parameter values and the A > I > E > O order. We assumed that A, I, E, and O were equally spaced on the confidence scale and so assigned them the integer

the same logit–logit analyses as above between PSYCOP's predictions and the overall data and obtained a comparable fit, $r(254) = .82$, rms error $=$ .45. Thus, PHM achieves as good a fit as PSYCOP. However, in fitting PSYCOP to the data, Rips used 9 parameters compared to our 5. Consequently, PHM provides a far more parsimonious account of these data than PSYCOP.

Mental Models

Mental models theory (MMT) proposes that syllogistic reasoning involves manipulating arbitrary mental exemplars which match the quantifiers. Consider the AA4 syllogism, which has a valid A conclusion. The first premise creates the following model (using Johnson-Laird's, 1983, notation):

$$\begin{array}{rcl} X & = & Y \\ X & = & Y \\ & & (Y) \end{array}$$

The X's represent objects that are X and the Y's objects that are Y. The ‘‘=’’ indicates that these are the same objects. The bracketed ‘‘Y’’ indicates an object that may or may not exist which is not an X. The second premise *All Y are Z* combines with this representation to yield:

$$\begin{array}{rcrcl} X & = & Y & = & Z \\ X & = & Y & = & Z \\ & & (Y) & = & Z \\ & & & & (Z) \end{array}$$

The conclusion that *All X are Z* can be ‘‘read off’’ this final model. Importantly, there is no other arrangement of these exemplars consistent with the premises—AA4 is a ‘‘one-model’’ syllogism. However, other syllogisms allow the construction of up to three mental models. Only conclusions that can be read off all models of the premises are valid. If no conclusion holds in all models, then the syllogism is invalid. Note that MMT embodies existential presupposition at the outset (following Aristotle), because it represents quantified statements by specific exemplars.

MMT correctly predicts that syllogism difficulty correlates with the num-

---

values 4, 3, 2, and 1, respectively. The reestimated parameter values were, $p_A = 32.85$, $p_I = 26.52$, $p_E = 20.18$, $p_O = 13.84$. Using these parameters, we still obtained as good a logit–logit fit to Rips' data as PSYCOP, $r(254) = .81$, rms residual $= .38$.

Note also that overfitting is not a possibility for the models reported above, because they have a very small number of parameters relative to the number of data points modeled.

ber of models that must be constructed. It also provides an explanation for the figural effect, in terms making the middle term (Y) contiguous in a mental model (Johnson-Laird & Byrne, 1991). Thus, in AA4, the Y terms in the second premise ‘‘naturally’’ combine to form an integrated model, where X enters working memory first followed by Z. By contrast, forming an integrated model for AA1 (All Y are X, All Z are Y) involves constructing the model for the second premise first and then integrating it with a model for the first premise. This leads to a model where Z enters first followed by X. Hence, if memory is first-in first-out, figure 1 syllogisms lead to X–Z conclusions and figure 4 syllogisms lead to Z–X conclusions, as observed in the data. We now argue that PHM may provide a better explanation of these data.

Single and multiple model syllogisms. For the valid syllogisms, the distinction between one, two, and three model syllogisms in MMT maps directly onto distinctions in PHM (see Appendix C). All one model syllogisms have *informative* (AIE) conclusions. Multiple model syllogisms have an uninformative (O) conclusion. Moreover, the distinction between two and three model syllogisms mirrors the distinction between syllogisms for which the *min*-heuristic does (AO and OA: 2 model syllogisms) or does not (IE and EI: 3 model syllogisms) work. Thus, PHM captures the distinctions in difficulty between the valid syllogistic inferences as well as MMT.

EI and ∃ p syllogisms. The EI syllogisms and the ∃p syllogisms provide a test of MMT. Apart from AA3, these are all 3 model syllogisms and are, hence, equally difficult according to MMT. But empirically participants draw more valid inferences for the EI syllogisms. We performed a meta-analysis for just these syllogisms, testing the association between syllogism type and validity. For each study, we tested the association using Fisher’s exact test in the direction of association present. We assigned a positive or negative z score to each result, setting z to 0 if the test yielded $p > .5$ in either direction (see Pollard, 1985, and Oaksford & Chater, 1994). We calculated a combined z score using Stouffer’s method (Wolf, 1986). There was a highly significant crossover interaction: for the EI syllogisms, participants preferred the logically valid conclusion, whereas for the ∃p syllogisms, they preferred the logically invalid conclusion ($z = 3.43$, $p < .0001$).

Thus, it appears that people do not make existential presuppositions, violating the predictions of MMT.[15] PHM does better here than mental models. Although the above interaction is highly significant, recall that there were no significant differences for the EI syllogisms. In contrast, there were sig-

[15] This result is harder to explain for mental models because in J-LS and J-LB the syllogistic premises were presented using the definite article in the subject term, e.g., All *the* Artists are Beekeepers. The motivation was to unambiguously signal that an existential presupposition was being made. However, despite this precaution, participants are clearly more inclined to follow the *min*-heuristic.

nificantly more responses predicted by PHM than by validity for the ∃p syllogisms (see section on *Logical Validity*). So, while mental models must explain away the significant effects for the ∃p syllogisms, PHM only has to allow that for EI syllogisms there is a nonsignificant tendency to prefer the conclusion type that coincides with validity.

The NVC response. According to MMT, participants should respond NVC if no conclusion is found to be true in all possible models. Erroneous NVC responses arise when they do not construct all the possible models. Thus, MMT predicts that participants should be equally likely to make the NVC response for all syllogisms requiring the same number of models. We focus on two model syllogisms, for which each *max*-premise type occurs. PHM predicts that *max*-premise type will affect NVC responses, whereas MMT predicts no difference. We performed a linear contrast on the NVC response as in the section *No Valid Conclusion*, but only for the two model syllogisms. There was a highly significant linear contrast in the predicted direction, $F(1, 12) = 109.43$, $MS_e = 55.30$, $p < .0001$. These results favor PHM over MMT.

Effects of figure on syllogism difficulty. Johnson-Laird and Bara (1984) argued that MMT predicts the order $4 > 1 > 2 > 3$ over the syllogistic figures in the ease with which participants draw valid inferences. But a meta-analysis fails to confirm this prediction. A one-way ANOVA with the mean percentage of the logically correct conclusion type as the dependent variable and with figure as a within studies factor revealed no significant differences between figures, $F(3, 12) = 1.10$, $MS_e = 73.62$, $p = .39$. Thus, a key prediction of MMT is not empirically supported in our meta-analysis and, therefore, PHM requires no modification to account for putative effects of figure on syllogism difficulty.

We conclude that PHM captures the data explained by MMT, but also provides a more accurate and detailed account of that data.

Deduction as Verbal Reasoning

Polk and Newell (1988, 1995) account for syllogistic reasoning within their framework of ‘‘deduction as verbal reasoning’’ (DVR). We show below that PHM captures all the aggregate data that DVR explains, except belief bias effects (discussed below). But DVR also successfully models individual subject data, which presents an important challenge for future work with PHM.

Polk and Newell derive 14 predictions from DVR. First, DVR predicts more NVC responses when participants are placed under time pressure. This prediction is not specific to DVR, but applies to any process model where deciding on the appropriate response is time-consuming. Under time pressure, these operations may not be completed, causing an NVC response. Second, DVR predicts that errors on valid syllogisms (excluding NVC responses) are typically logically *consistent* with the premises. PHM also predicts this, because errors arise from the use of implicatures, which are always logically consistent with the valid conclusion.

DVR also predicts more NVC in Figs. 2 and 3 than in Figs. 1 and 4. Polk and Newell consider four pairwise comparisons (1 vs 2, 1 vs 3, 4 vs 2, and 4 vs 3), for all 64 syllogisms, for the 27 Aristotelian valid syllogisms, and for the 37 invalid syllogisms, giving 12 predictions in all. Polk and Newell do not report statistics for these predictions on the data they consider. We therefore tested these meta-analytically using the three production tasks considered above.

For all 64 syllogisms, two of the four predictions (4 vs 2 and 4 vs 3) were significant (1 vs 2, $t(30) = 1.41$, $p = .09$; 1 vs 3, $t(30) = .90$, $p = .19$; 4 vs 2, $t(30) = 2.73$, $p < .025$; 4 vs 3, $t(30) = 1.75$, $p < .05$). But this mixed picture becomes clearer when valid and invalid syllogisms are considered separately. For the valid syllogisms, *no* comparisons were significant (1 vs 2, $t(10) = .29$, $p = .39$; 1 vs 3, $t(13) = .39$, $p = .35$; 4 vs 2, $t(10) = .63$, $p = .27$; 4 vs 3, $t(13) = .78$, $p = .23$), whereas for the invalid syllogisms, *all* comparisons were significant (1 vs 2, $t(18) = 1.74$, $p < .05$; 1 vs 3, $t(15) = 2.88$, $p < .01$; 4 vs 2, $t(18) = 3.23$, $p < .0025$; 4 vs 3, $t(15) = 5.90$, $p < .0001$). Summing up, only six of the 12 predictions are confirmed, providing equivocal support for DVR.

PHM, by contrast, can explain these effects. The poor performance on figures 2 and 3 arises because these figures contain all eight syllogisms where the *attachment*-heuristic fails (AA2, AA3, II2, II3, EE2, EE3, OO2, and OO3). If attachment fails, participants may balk at attempting to produce a conclusion and, hence, respond NVC. All these syllogisms are invalid, correctly predicting that differences between figures will occur only for the invalid syllogisms.

To test this prediction, we divided the syllogisms into ‘‘doubles’’ (the 16 syllogisms where both premises have the same logical type) and ‘‘singles’’ (the other 48 syllogisms). All syllogisms where attachment fails are doubles. Thus, if attachment explains figural differences in NVC responses, then those differences should occur for the doubles but not for the singles. To test this, we conducted a 2 × 2 ANOVA with premise-type (doubles vs singles) and figure (1 and 4 vs 2 and 3) as between syllogisms factors and percentage of NVC responses as the dependent variable. There was a significant main effect of figure, $F(1, 60) = 8.23$, $MS_e = 539.44$, $p < .01$, with more NVC responses for Figs. 2 and 3 than for Figs. 1 and 4, as both DVR and PHM predict. However, there was also a significant main effect of premise-type, $F(1, 60) = 9.02$, $MS_e = 539.44$, $p < .005$, with more NVC responses for doubles than for singles. Although the interaction was not significant, $F(1, 60) = 2.38$, $MS_e = 539.44$, $p = .128$, simple effects comparisons at each level of premise-type, revealed significantly more NVC responses for Figs. 2 and 3 than for Figs. 1 and 4 for the doubles, $F(1, 60) = 6.50$, $MS_e =$

TABLE 5
Predictions of the *min*-Heuristic, the Atmosphere Hypothesis, and Matching on the AIEO Syllogisms

Syllogism	*min*-Heuristic	Atmosphere	Matching
AA	**A,** I	**A**	**A**
AI	**I,** O	**I**	**I**
AE	**E,** O	**E**	**E**
AO	**O,** I	**O**	**O**
II	**I,** O	**I**	**I**
IE	**E,** O	**O**	**E**
IO	**O,** I	**O**	**I, O**
EE	**E,** O	**E**	**E**
EO	**O,** E, I	**O**	**E**
OO	**O,** I	**O**	**O**

Note. The *min*-heuristic column shows the predictions of the *min*-heuristic; the atmosphere column shows the predictions of the atmosphere hypothesis; and the matching column shows the predictions of matching. The letters not in bold face represent the *p*-entailments in PHM.

539.44, $p < .025$, but *not* for the singles, $F(1, 60) = 1.76$, $MS_e = 539.44$, $p = .190$). This suggests that the overall main effect of figure is largely due to the difference between figures for the doubles alone. To confirm this, we reanalyzed the overall comparisons with the 16 doubles removed, and, as PHM predicts, we found no significant differences (1 vs 2, $t(26) = .54$, $p = .30$; 1 vs 3, $t(26) = .11$, $p = .46$; 4 vs 2, $t(26) = .15$, $p < .08$; 4 vs 3, $t(26) = .92$, $p = .18$). Thus, PHM would appear to provide a more accurate account of the effects of figure than DVR.

Summing up, PHM captures the aggregate effects captured by DVR. Comparing PHM with DVR's models of individual data is an area for future research.

Atmosphere, Matching, and Conversion

Atmosphere, matching, and conversion are nonlogical accounts of syllogistic reasoning, with similar predictions to PHM.

The atmosphere hypothesis (Woodworth & Sells, 1935; Begg & Denny, 1969) has two parts: *the principle of quality*, that if one or both premises are negative (E or O), the conclusion should be negative; otherwise, it is positive (A or I); and *the principle of quantity*, that if one or both premises are particular (I or O), then the conclusion will be particular; otherwise, it is universal (A or E). Although its predictions (Table 5) are well confirmed experimentally (Begg & Denny, 1969; Jackson, 1982), atmosphere has attracted relatively little attention, presumably because it seems theoretically unmotivated.

PHM explains why atmosphere provides good data fits: The predictions of atmosphere and the *min*-heuristic are almost identical. They only disagree for the EI syllogisms, where atmosphere predicts an O conclusion and *min* predicts an E conclusion with an O conclusion as a *p*-entailment. In the meta-analysis reported above (see section on Logical Validity), there were no significant differences between the frequencies with which E and O conclusions were drawn for the EI syllogisms, although the trend favors atmosphere.

PHM is much more comprehensive than atmosphere. For example, PHM explains the second most frequently chosen conclusion (in terms of the *p*-entailments), where atmosphere makes no predictions. This is particularly serious for Rips's (1994) data, where these responses are as frequent as those predicted by the *min*-heuristic. Moreover, atmosphere does not apply to the generalized quantifiers, *Most* and *Few*, whereas PHM generalizes straightforwardly to these cases.

Matching (Wetherick, 1989, 1993; Wetherick & Gilhooly, 1990) resembles the *min*-heuristic, but relies on ''conservatism'' rather than informativeness. A quantified statement is *conservative* to the extent that it commits the speaker to a small number of objects. Thus, E-statements are the most conservative because they do not presuppose any objects, A-statements are the least conservative, and I- and O-statements are in between, i.e.,: $C(A) < C(I) = C(O) < C(E)$. The matching hypothesis is that people select conclusions of the same type as the *most* conservative premise. Matching (Table 5) gives good empirical fits (Wetherick, 1993).

Some predictions of matching conflict with PHM. For IO syllogisms, matching predicts no differences between I and O conclusions because conservatism does not differentiate between I and O, whereas PHM, via the *min*-heuristic, predicts more O responses; for EO syllogisms, matching predicts a preference for E over O conclusions, whereas PHM, again via *min*, predicts the opposite. To test these predictions, we performed Fisher's combined tests, as used above to test whether people drew the *p*-entailments. As PHM predicts, for most IO syllogisms, participants prefer O to I ($\chi^2(10) = 51.29$, $p < .0001$). Moreover, participants prefer the O to E for the EO syllogisms, although this difference is not quite significant ($\chi^2(10) = 15.25$, $.05 < p < .1$). Thus, where matching and PHM diverge, the evidence favors PHM.

The comparison for the IO syllogism assumes that matching predicts no differences between I- and O-conclusions. Can matching be saved by assuming that O-statements are more conservative that I-statements? *Logically*, this seems unjustified, as both require the existence of only one entity. However, Newstead, Pollard, and Riezebos (1987) suggest there is a *psychological* difference between the two: People treat I-statements as presupposing fewer entities than O-statements (i.e., I is more conservative) (Evans, Newstead, & Byrne, 1993). Unfortunately, the opposite is required for matching to explain the experimental data. Finally, as for atmosphere, matching does not apply to the generalized quantifiers, which we explore below.

Conversion. A classic explanation of syllogistic reasoning errors is that people erroneously ''convert'' one or more premises, e.g., representing *All A are B* as *All B are A* (e.g., Ceraso & Provitera, 1971; Chapman & Chapman, 1959; Revlis, 1975a, 1975b) and then draw a logically valid conclusion from these incorrect premises. Conversion allows, for example, an AA2 syllogism to be ''converted'' into an AA1, AA3, and AA4—thus conversion effectively ignores figure. Thus, if any syllogism with the given premise types has a valid conclusion, then it is assumed to be a possible response, whatever the figure.

PHM captures the predictions of conversion, but also accounts for many additional effects. For example, PHM correctly predicts the conclusions for syllogisms which are invalid in every figure (IO, OI, EE, EO, OE, OO) by the *p*-entailments, where conversion does not apply. This is particularly problematic for conversion in Rips's (1994) evaluation task, where responses for conclusions licensed by PHM, but not conversion, are as frequent as responses predicted by both accounts.

In conclusion, atmosphere, matching, and conversion succeed because they coarsely approximate PHM. But PHM would appear preferable to each because it gives a more comprehensive account of the data and has greater theoretical motivation.

EXPERIMENTS WITH GENERALIZED QUANTIFIERS

In this section, we report two experiments using the generalized quantifiers Most and Few. These experiments test the novel predictions of PHM which appear to be beyond the scope of other current theories of syllogistic reasoning, which are currently limited to the standard logical quantifiers. These data provide a challenge for mental models and mental logic accounts to develop reasoning algorithms that account for the pattern of performance we observe.

Experiment 1: The AMFO Quantifiers

According to PHM, participants' behavior on syllogistic reasoning tasks is not determined by logical reasoning, but by rational informational heuristics. Using the standard AIEO materials, this seems reasonable, because our rational heuristics agree almost perfectly with standard logic for informative, valid inferences.

By adding M and F to the standard AIEO quantifiers, we can extend the set of possible syllogistic reasoning problems. With 6 quantifier types, the first premise can take 6 values, as can the second. Each of these 36 possibilities can be in each of the 4 figures, giving a total of 144 syllogisms. Pilot work suggested that it was too taxing for participants to complete all 144 syllogisms so, in order to obtain reliable data, we gave participants a set of

64 syllogisms, using only the quantifiers AMFO. In Experiment 2, we used the quantifiers MFIE.

PHM generates predictions for these experiments in exactly the same way as for the standard AIEO syllogisms. In contrast, logic based accounts, including mental logics and mental models, currently make no predictions for these data.

Method

Participants. Twenty participants from the University of Warwick Participant Panel took part in these experiments. Each was paid £3.00 for taking part. None had any prior experience of psychological reasoning tasks.

Design. Each participant received a booklet containing the same set of 64 syllogisms, in different random orders.

Materials. The materials consisted of a booklet with an instruction page and 64 syllogisms involving the quantifier types *All* (A), *Most* (M), *Few* (F), and *Some-Not* (O). As in the standard task, the response options were the same as the possible premise types. One consequence is that the OO2 syllogism is treated as invalid in our subsequent analyses because the *p*-valid I conclusion was not an available response. Thus, for this experiment, 38 syllogisms were treated as *p*-valid (rather than 39). Following Dickstein, we presented the conclusions in a standard Z–X order, rather than including all possible response options.

All the syllogisms involved lexical categories that have been standardly used in syllogistic reasoning tasks, e.g., *Most artists are beekeepers; Few chemists are beekeepers.*

Procedure. Participants were tested individually. The front page of the instructions read:

> This experiment looks at how people reason with simple sentences. There are 64 questions and each question consists of 2 premises and 4 possible conclusions. Your task is to tick the box next to each conclusion to indicate which of the four conclusions you think follow from the premises. You can tick more than one box if you think that more than one of the conclusions follows from the premises. If you think that *no* conclusion follows, then simply leave all the boxes blank.
> Thank you.

After reading the instructions, participants had to solve the 64 syllogisms. They were told that there was no time limit (typically, they completed the task in under one hour). Allowing participants to make as many responses as they think appropriate creates four independent response categories. This allows us to treat response category as an independent factor in subsequent data analysis.

Results and Discussion

Appendix D presents the results of Experiment 1. The *min*-heuristic predicts the modal response for each type of syllogism (in accordance with the *min*-heuristic column in Table 3). In contrast with the standard AEIO syllogisms, participants were very confident in the *min*-conclusion for all types of syllogism. This confirms that the *min*-heuristic is robust across quantifier types. This is striking because few (6 of the 64 syllogisms) of the conclusions recommended by the *min*-heuristic are logically valid. Indeed, confidence in the *min*-conclusion was as high for the only *p*-valid OO syllogisms as for the logically valid AA syllogisms. Further, confidence in the O conclusion for the OO syllogisms (68.75%) was higher than in the context of the

AEIO quantifiers (18.04%), suggesting some form of global context effect. This contextual effect is a puzzling and challenging result for any current theory of syllogisms. PHM accounts for 91.64% of participants' responses (ignoring the NVC response).

Participants were allowed to make more than one response. Consequently, *p*-entailment conclusions should also be observed, although not as frequently as in Rips (1994), where response competition is eliminated altogether. Appendix D reveals that, as PHM predicts, there were more *min*-heuristic responses that do not rely on *p*-entailments than those that do, and there were more *min*-heuristic responses that rely on *p*-entailments than responses that the *min*-heuristic does not predict. In sum, PHM seems to capture the majority of the effects in the data.

Modeling. We assess the overall fit between data and model in the same way as in our meta-analysis. Because confidence was high for all *max*-premise types (although the *max*-heuristic is in evidence; see below), we use a single parameter p_{min}(67.03) to capture the percentage of choices determined by the *min*-heuristic. With these premises, there are two kinds of *p*-entailment: one where a single entailment follows and one where two entailments follow more weakly. Therefore, we use p_{ent1}(38.93) to capture the percentage of choices determined by the *min*-heuristic, where a single *p*-entailment is allowed, and p_{ent2}(17.01) to capture the same percentage where two *p*-entailments are allowed (we assume that each entailment is equally probable). Including p_{error}(5.27), we only require four parameters in all. Overall fits involving all four responses, A, M, F and O, for all 64 syllogisms (256 data points in all), yielded a close fit between data and model, $r(254) = .94$ (rms error = .16). Thus, the model accounts for over 88% of the variance in the overall data. As PHM predicts, this pattern of results is very similar to the AIEO syllogisms. However, the results of this experiment can not be explained by logic based theories such as mental logic and mental models.

P-valid syllogisms. For the *p*-valid syllogisms, there were no significant differences between informative syllogisms—those with an A, M, or F conclusion—and the uninformative syllogisms—those with an O conclusion—where the *min*-heuristic applies (O_{min}), either by participants, $t(19) = .29$, ns (means: AMF = 67.81 (*SD:* 31.24); O_{min} = 65.28 (*SD:* 30.53)), or by materials, $t(32) = .61$, p = ns (means: AMF = 67.81 (*SD:* 11.69); O_{min} = 65.28 (*SD:* 12.54)). We interpret this result to indicate that, in the context of the unfamiliar syllogisms using generalized quantifiers, participants rely more heavily on the *min*-heuristic than for the AIEO syllogisms (see above). However, there were significant differences between the O_{min} syllogisms and those where the *min*-heuristic does not apply (O), i.e., AM2, MA2, MF1, FM4, both by participants, $t(19) = 4.31$, $p < .0005$ (means: O_{min} = 65.28 (*SD:* 30.53); O = 31.25 (*SD:* 38.79)) and by materials, $t(20)$

$= 5.26$, $p < .0001$ (means: $O_{min} = 65.28$ (*SD:* 12.54); $O = 31.25$ (*SD:* 4.79)).

P-invalid syllogisms. As in our meta-analysis, we excluded the *p*-valid syllogisms and recomputed the fit between data and model. The fit was as good as for the overall data, $r(102) = .92$ (rms error = .18). Consequently, for the AMFO syllogisms in Experiment 1, the *min*-heuristic accounts as well for the *p*-invalid as for the *p*-valid syllogisms.

The max-*heuristic.* The means and standard deviations for each *max*-premise type were: A = 71.07 (*SD:* 27.78), M = 64.25 (*SD:* 31.55), F = 54.58 (*SD:* 26.28), and O = 70.00 (*SD:* 33.05). Confidence was as high for the O *max*-premise type as for the A *max*-premise type (see above), which is not predicted by our account. Note, however, that there are only four syllogisms contributing to this mean confidence level for O conlusions. If these are removed, then the linear trend is statistically significant.

P-validity. To test for a *p*-validity effect, we compared the mean percentage of *min*-responses for the *p*-valid and *p*-invalid syllogisms. There were no significant differences, either by participants, $t(19) = .79$, ns (means: *p*-valid = 62.76 (*SD:* 22.05); *p*-invalid = 65.38 (*SD:* 27.59)) or by materials, $t(62) = .75$, ns (means: *p*-valid = 62.76 (*SD:* 11.87); *p*-invalid = 65.38 (*SD:* 12.19)). Consequently, it would appear that participants follow the *min*-heuristic as diligently for the *p*-invalid as for the *p*-valid syllogisms.

We checked whether response availability (recall that only the Z–X conclusion order was available) had any effect on the *min*-response by comparing the frequency of *min*-responses for *p*-valid syllogisms where the *p*-valid response was available with the frequency of *min*-responses for *p*-valid syllogisms where it was not available. There were no significant differences between these two groups, either by participants, $t(19) = .75$, $p = .46$ (means: available = 70.46 (*SD:* 22.63); unavailable = 68.18 (*SD:* 25.97)) or by materials, $t(36) = .42$, $p = .68$ (means: available = 70.46 (*SD:* 10.83); unavailable = 68.18 (*SD:* 14.37)). Consequently, conclusion order has little influence on participants' responses and so, as PHM uniquely predicts, conclusion type can be arrived at without fixing conclusion order (although not vice versa).

The min-*heuristic and the* p-*entailments.* We evaluated these predictions in a one-way, within participants ANOVA with four levels corresponding to the different response categories: *min*-heuristic without *p*-entailments (MIN), the *min*-heuristic plus O *p*-entailments (MIN_O), the *min*-heuristic plus M or F *p*-entailments ($MIN_{m \vee f}$), and responses not predicted by the *min*-heuristic (NON-MIN). The dependent variable was the percentage of these responses made. We assessed all the pairwise predictions of the model using planned contrasts. All were significant in the predicted direction: MIN > MIN_O, $F(1, 57) = 18.05$, $MS_e = 493.28$, $p < .0001$ (means: MIN = 65.78, *SD:* 25.28; $MIN_O = 35.94$, *SD:* 34.02); $MIN_O > MIN_{m \vee f}$, $F(1, 57) = 5.27$, $MS_e = 493.28$, $p < .05$ (mean: $MIN_{m \vee f} = 19.82$, *SD:* 15.92); and $MIN_{m \vee f}$

$>$ NON-MIN, $F(1, 57) = 4.37$, $MS_e = 493.28$, $p < .05$ (mean: NON-MIN $= 5.14$, *SD:* 6.79). In summary, these analyses demonstrate that participants are drawing the predicted *p*-entailments, the existence of which is one of the major novel predictions of our probabilistic analysis.

Figure and task difficulty. Does figure affect task difficulty, as mental models (MMT) predicts, for syllogisms that use generalized quantifiers and *p*-validity? Using the fixed response order meant that half of the 38 *p*-valid syllogisms had the *p*-valid response option available. For figure 1, all the 13 *p*-valid syllogisms had the *p*-valid conclusion available; for figure 2, there were three out of six; for figure 3, there were three out of six; but none of the 13 *p*-valid figure 4 syllogisms had the *p*-valid conclusion available. We therefore performed a one-way ANOVA with figure as a within participants factor (excluding figure 4) and percentage of *p*-valid conclusions drawn as the dependent variable. The linear trend predicted by mental models was not observed, $F(1, 38) = 1.11$, $MS_e = 447.04$, ns. Moreover, if anything, the trend was in the opposite direction to that predicted by mental models (means: one: 64.62(23.73), two: 65.00(31.48), three: 71.67(34.67)). We also failed to discover any effects of figure when including more of the data by performing similar analyses using (i) all *p*-valid syllogisms and (ii) all syllogisms, with the proportion of *min*-responses as the dependent variable. We therefore conclude that, contrary to MMT, figure has no effects on the ease of drawing syllogistic conclusions with generalized quantifiers.

Learning effects? It could be argued that PHM does not reflect people's normal reasoning strategies but reflects learned strategies acquired in response to the demands of the task. If so, then the generality of PHM would be greatly reduced. Of course, this would also question the wider relevance of studying syllogisms for understanding real human reasoning. Nonetheless, if participants were learning these strategies, then we would expect more responses that accord with the *min*-heuristic as the experiment progresses. We therefore divided participants' responses into 8 trial blocks in the order in which participants attempted to solve the 64 syllogisms. Each block consisted of eight syllogisms. If there is a learning effect, then we would expect a significant linear trend in the number of *min*-conclusions over these trial blocks, depending on their position in the trial sequence. No such trend was observed, $F(1, 133) < 1$. It would appear, therefore, that participants bring to these tasks strategies that are based on their normal reasoning behavior and they are not simply developing strategies on-line in response to task demands.

Summary

Presenting participants with syllogistic arguments using the quantifiers AMFO reveals the pattern of results predicted by PHM. Indeed, the results are very similar to those obtained in standard syllogistic reasoning tasks. This indicates that other theories should treat reasoning with generalized quantifiers uniformly with reasoning with the standard quantifiers, as first suggested by Johnson-Laird (1983). These data therefore provides a challenge for mental model and mental logic accounts of syllogistic reasoning.

Experiment 2: The MFIE Quantifiers

This experiment was conducted in the same way as Experiment 1 but using MFI and E. The only significant difference that PHM predicts between this experiment and Experiment 1 is that strategies based on our informational order might be expected to be weaker. This is because these four premise types are adjacent in the informational order, rather than being spread out along the entire continuum, as in AMFO.

Method

Participants. Twenty participants from the University of Warwick Participant Panel took part in these experiments. Each participant was paid £3.00 for taking part. No participants had any prior experience of psychological reasoning tasks.

Design. The design was the same as in Experiment 1.

Materials. Other than the change of quantifiers, the materials were the same as in Experiment 1. Response options were limited to those quantifiers contained in the premises as in Experiment 1. One consequence is that the MF1 and FF1 syllogisms are treated as invalid in our subsequent analyses because the *p*-valid O conclusion was not an available response. Thus, for this experiment, 24 syllogisms were treated as *p*-valid (rather than 26).

Procedure. The procedure was the same as in Experiment 1.

Results and Discussion

The results of Experiment 2 are shown in Appendix E. The *min*-heuristic predicts the modal response for most syllogism types (in accordance with the *min*-heuristic column in Table 3). Moreover, as predicted, in the midrange of informational values represented by the MFIE syllogisms, participants are less confident in the min conclusion for all types of syllogism. This finding only makes sense from the perspective of the *min*-heuristic which relies on the ability to discriminate between the informational values of the premises. PHM accounts for 65.74% of participants' responses (ignoring cases where participants choose no conclusion, i.e., NVC). This is less than for the AMFO syllogisms. Again, there seems to be some form of global context effect as we discussed in Experiment 1 for the OO syllogisms. A similar effect was observed here, such that the MM, MF, and FF syllogisms were endorsed more strongly in Experiment 1 than in Experiment 2. Nonetheless, the *min*-heuristic is clearly in evidence.

The *p*-entailments in PHM predict: (i) that there should be more *min*-heuristic responses that don't rely on *p*-entailments than do rely on *p*-entailments and (ii) that there should be more *min*-heuristic responses that rely on *p*-entailments than responses that the *min*-heuristic does not predict. Casual observation of Appendix E provides some support for this prediction. However, the data are less clear-cut than for the AMFO data. In sum, Appendix E reveals that our simple informational strategy is in evidence in the MFIE data. Before turning to a more fine-grained analysis, we assess quantitatively how much of the data PHM captures.

Modeling. To assess the overall fit between data and model, we estimated the same parameters as for Experiment 1 (p_{min} = 37.03, p_{ent1} = 16.00, p_{ent2} = 14.13, p_{error} = 12.28). The product moment correlation between the 256 observed and predicted values was $r(254) = .65$, and the root-mean-square deviation was .26. Thus, the model accounts for 42% of the variance in the overall data. The fit for the MFIE data is less good than for the AMFO data. However, recall that these analyses serve a comparative function. Rips (1994) achieved a fit of $r(254) = .74$ to his own data but only by using *ten* parameters. Thus, the level of fit we obtain even for the MFIE data is comparable to that of other theoretical approaches that can not even be applied to these data.

P-valid syllogisms. Because the uninformative O conclusion was not available as a response category, analyses comparing the informative and uninformative *p*-valid syllogisms could not be carried out for Experiment 2.

P-invalid syllogisms. Excluding the *p*-valid syllogisms still yielded a fit as good as for the overall data, $r(164) = .65$ (rms error = .25). Consequently, for the MFIE syllogisms in Experiment 2, the *min*-heuristic accounts as well for the *p*-invalid as for the *p*-valid syllogisms.

The max-*heuristic.* The means and standard deviations for each *max*-premise type were: M = 42.50 (*SD:* 33.92), F = 37.08 (*SD:* 25.28), I = 31.25 (*SD:* 31.93), and E = 31.25 (*SD:* 43.59). There was a close to significant linear trend in the predicted order, $F(1, 57) = 3.36$, $MS_e = 465.77$, $p = .072$. The lack of significance is because the MFIE quantifiers are more closely bunched along the informational ordering than the AMFO quantifiers. Consequently, although there is some evidence for the *max*-heuristic in these data, the effects are not as strong as we observed for the AMFO syllogisms.

P-validity. To test for a *p*-validity effect, we performed the same analyses as for Experiment 1. In contrast to Experiment 1, there were significantly more *min*-responses for the *p*-valid than for the *p*-invalid syllogisms. This difference was reliable both by participants, $t(19) = 2.07$, $p < .05$ (one-tailed) (means: *p*-valid = 40.21 (*SD:* 30.33); *p*-invalid = 35.12 (*SD:* 28.68)), and, albeit marginally, by materials, $t(62) = 1.61$, $p = .056$ (one-tailed) (means: *p*-valid = 40.21 (*SD:* 12.89); *p*-invalid = 35.12 (*SD:* 11.79)). The concept of *p*-validity and its applicability to syllogistic reasoning with generalized quantifiers receives some support from these findings. Although participants' behavior is generally guided by the *min*-heuristic, they are still sensitive to *p*-validity.

We again checked whether response availability had any effect on participants' use of the *min*-heuristic in Experiment 2 by comparing the frequency of *min*-responses for *p*-valid syllogisms when the *p*-valid response was available and when it was not available (the EI syllogisms and MM4 and MF4). There were significant differences between these two groups, both by participants, $t(19) = 2.07$, $p < .05$ (means: available = 46.43 (*SD:* 35.34); unavailable = 31.50 (*SD:* 32.65)) and by materials, $t(22) = 3.37$, $p < .005$ (means: available = 46.43 (*SD:* 11.51); unavailable = 31.50 (*SD:* 9.44)). However, the point of these analyses was to see if the fixed Z–X conclusion order affected participants' choice of conclusion type. But the appropriate *p*-valid response was unavailable for the EI syllogisms only because, like all other syllogistic reasoning experiments, we restricted the range of response options to the quantifiers used in the premises. Therefore, to more accurately assess the effects of conclusion order, we excluded the EI syllogisms from the analysis. When this was done, there were no significant differences either by participants, $t(19) = .18$, $p = .86$ (means: available = 46.43 (*SD:* 35.34); unavailable = 47.50 (*SD:* 44.35)) or by materials, $t(14) = .13$, $p = .90$ (means: available = 46.43 (*SD:* 11.51); unavailable = 47.50 (*SD:* 3.54)). Consequently, as for the AMFO syllogisms, conclusion order does not influence choice of conclusion type and so, as PHM uniquely predicts, conclusion type can be arrived at without fixing conclusion order (although not vice versa).

The min*-heuristic and the* p*-entailments.* We assessed these predictions in the same way as in Experiment 1. The only difference was that a MIN_I response category (the *min*-heuristic plus I *p*-entailments) replaces MIN_O. In planned contrasts testing for all pairwise difference, only the first was significant: $\text{MIN} > \text{MIN}_\text{I}$, $F(1, 57) = 12.72$, $MS_e = 319.36$, $p < .001$ (means: MIN = 37.03, *SD:* 28.82; MIN_I = 16.88, *SD:* 25.58). No other comparison reached significance. However, PHM does predict a linear trend such that $\text{MIN} > \text{MIN}_\text{I} > \text{MIN}_{\text{m}\vee\text{f}} > \text{NON-MIN}$, which was significant, $F(1, 57) = 18.68$, $MS_e = 319.36$, $p < .0001$, and is consistent with participants drawing the predicted *p*-entailments. In sum, these data offer strong support for the *min*-heuristic and some support for the *p*-entailments.

Figure and task difficulty. We assessed whether figure affects task difficulty, as predicted by MMT, in the same way as in Experiment 1. For figures 1 to 3, *p*-valid syllogisms had the *p*-valid conclusion available. For figure 4, there were five out of seven available. Because the *p*-valid O conclusion was not available, the EI syllogisms were excluded (two for each figure). A one-way ANOVA with figure as a within-subjects factor and percentage of *p*-valid conclusions drawn as the dependent variable was not significant, $F(3, 57) = 2.25$, $MS_e = 406.90$, $p = .09$ (means: four: 50.00(41.18), one: 50.00(36.42), two: 55.00(51.04), three: 39.00(25.82)). In contrast to Experiment 1, using all 64 syllogisms with the percentage of *min*-responses as the dependent variable revealed a significant trend in the direction predicted by MMT, $F(1, 57) = 5.35$, $MS_e = 65.76$, $p < .05$ (means: four: 39.69(30.90), one: 38.13(29.17), two: 36.25(27.77), three: 34.06(47.23)). However, the existence of figural effects for the *valid* syllogisms is important for MMT because they are supposed to arise in drawing logically valid conclusions. Consequently, MMT must predict stronger effects for the (p-) valid than for the (p-) invalid syllogisms. However, these effects are only evident when looking at the overall data. Therefore, the figural effect observed here can not be interpreted as arising from the process of computing p-valid inferences.

Learning effects? As in Experiment 1, we checked for learning effects by looking for a trend in the number of *min*-conclusions over trial blocks depending on their position in the trial sequence. No such trend was observed, $F(1, 133) < 1$. These results further strengthen our conclusion that participants bring strategies to these tasks that are based on their normal reasoning behavior and that they are not simply developing strategies on-line in response to the demands of the task.

Summary

Experiment 2 further confirms PHM for data that falls outside the scope of other accounts. Notice that having conducted both Experiments 1 and 2, as well as a meta-analysis over the data from the standard syllogisms, we have shown that PHM applies successfully across all 144 possible syllogisms involving the quantifiers AMFIEO (although not with all possible conclusion types). The uniform pattern in these data across generalized and standard quantifiers challenges other theories of syllogistic reasoning, which currently apply only to the standard logical quantifiers, to generalize to these data.

GENERAL DISCUSSION

This paper presents a set of ‘‘fast and frugal’’ heuristics for syllogistic reasoning, which we show approximate a rational standard for reasoning with quantifiers under a probabilistic interpretation. These heuristics reliably *generate* the p-valid conclusion, if there is one (the *min*- and *attachment*-heuristics). They also provide a relatively crude measure to *test* whether a syllogism is valid or not (the *max*-heuristic). However, this means that people are likely to generate and propose conclusions for the invalid syllogisms as well. PHM accounts for this pattern of errors not only for the standard logical quantifiers but also for syllogisms involving the generalized quantifiers Most and Few. In this discussion, we first discuss the challenge posed to alternative theories of syllogistic reasoning by our extension to generalized quantifiers. This challenge is totally independent of the validity of PHM. However, we believe that PHM provides a well-motivated account of syllogistic reasoning. Therefore, second, we discuss some theoretical and empirical issues that might be thought to question the plausibility of PHM.

P-Validity, Mental Logic, and Mental Models

Aside from the validity of PHM, this paper makes two important contributions to the psychology of syllogistic reasoning. First, it defines a notion of validity applicable both to the standard quantifiers and to the generalized quantifiers, Most and Few. To generalize to these quantifiers, other theories of syllogistic reasoning must either adopt our notion of *p*-validity, or develop an alternative notion of validity. That is, they must define their own computational level theory of the inferences people should draw with generalized syllogisms or they must adopt ours. Only then can they differentiate correct from incorrect performance when people reason with these syllogisms. Second, we have conducted experiments using generalized quantifiers, which show a pattern of performance apparently uniform with performance using the standard quantifiers. People typically choose the correct conclusion if there is one, but are relatively poor at distinguishing *p*-valid from *p*-invalid conclusions. *Independently of* PHM, *all other theories of conditional reasoning must show how to generalize to this pattern of performance.*

Mental logic approaches seem to be in the worst shape in this respect, because they rely only on logical machinery that could not generalize to Most and Few. Mental models seems to be in better shape here because, as Johnson-Laird (1983, 1994) has urged, one main advantage of the mental models formalism is that it can naturally represent generalized quantifiers and, moreover, deal with certain kinds of probabilistic reasoning (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, in press). However, representing generalized quantifiers does not define a proof procedure for deriving conclusions from premises containing quantifiers like Most and Few. In particular, with these quantifiers, the guiding principle of mental models, that people search for counterexamples, will be of no use—with probabilistic relations counterexamples will always be possible and so finding one tells you little about the *p*-validity of an inference. Whatever the outcome, it is clear that mental logic and mental model theories must face the central challenge of accounting for generalized quantifiers and for the pattern of inferences revealed in Experiments 1 and 2. With respect to this challenge, PHM clearly has the advantage over the other main theories of syllogistic reasoning. However, it could be argued that, despite the generality of PHM within the domain of syllogisms, it is less plausible than other accounts that generalize to other forms of inference. In the next section, we consider some possible objections to PHM, to which we reply.

Objections and Replies

In this section, we tackle some possible theoretical and empirical objections to PHM that may question its plausibility as a general model of syllogistic reasoning.

Theoretical Objections

There are two possible sources of doubt concerning the plausibility of PHM. First, the heuristics we have proposed to explain syllogistic reasoning, although providing good fits to the existing data and generalizing naturally to generalized quantifiers, may seem very domain specific. This may be felt to contrast with other accounts like mental logics and mental models that, although failing to generalize to generalized quantifiers, have been shown to account for a broad range of data from other reasoning tasks. The second species of possible doubt about PHM concerns how people could have acquired these heuristics which we argue they bring to bear naturally on the syllogistic reasoning task. We begin by looking at the generality of the *min*- and *max*-heuristics.

Generality. The first thing to note is that similar strategies apply and are in evidence in other areas of reasoning. For example, an heuristic similar to the *max*-heuristic is used in well-known AI systems for medical diagnosis, such as MYCIN and EMYCIN, to compute the believability of conclusions (Gordon & Shortliffe, 1985; Shortliffe & Buchanan, 1990). Moreover, George (1997) argues that a similar heuristic to the *max*-heuristic is evident in human data on conditional inference with uncertain premises. Note that, given our probabilistic semantics, the conditional premise can be characterized as a quantified statement. George (1997) used three levels of probability: certain (C), very probable (VP), and not very probable (NVP), e.g., if A, then certainly B. According to our probabilistic semantics, we can interpret conditional premises containing these terms as corresponding to quantified premises using All, Most, and Few, respectively. According to probability theory, it has to be the case that the probability of the conclusion is, on average, higher for All than for Most and higher for Most than for Few. That is, on average, our confidence in a conclusion should be proportional to the informativeness of the conditional premise. So if you don't know the precise probabilities, then a good heuristic for how confident you should be in the conclusion is to follow the A $>$ M $>$ F order, as we have recommended in syllogistic reasoning. Moreover, George's (1997) Experiment 2 confirms that people conform to this average ordering: the average rated probability of a conclusion when the conditional premise contained ''certainly'' (All) was .70, when it contained ''very probable'' (Most) it was .60, and when it contained ''not very probable'' (Few) it was .36 (these expected values were calculated by rescaling George's (1997) rating scale into the 0–1 probability scale).

Acquisition. How would naive participants acquire the heuristics we propose in this paper? Although the *min*- and *max*-heuristics may have analogs in other areas of reasoning, we believe that their specific application in syllogistic reasoning is acquired from the ubiquity of various forms of implicit syllogistic reasoning that people must be carrying out all the time. We also believe that this is how the *attachment*-heuristic is learned and so, before outlining these forms of implicit syllogistic reasoning that we believe underpin our explicit abilities in syllogistic reasoning tasks, we first outline just how simple attachment is to master.

Attachment is the most task specific heuristic in PHM. However, on our assumption that *min* only generates a conclusion type, participants must have a way of making the simple binary choice concerning the ordering of end terms in the conclusion. The first point worth making is that attachment is a very simple pattern matching strategy that is very easy for participants to apply despite its somewhat cumbersome definition. The second point worth making is that attachment generally has an even simpler characterization: go with the order of end terms dictated by the premise most similar to the generated conclusion type, but if tied for similarity, leave the order of terms as they appear in the premises (this is very similar to the description of the figural effect in Stenning & Yule, 1997).[16] This works for the *min*-responses by definition. So, for example, the conclusion of AI1 is Some Z are X, because Z is in subject position in the I premise, whereas the conclusion of AI3 is Some X are Z, because Z is now in the predicate position in the I premise. This definition also works for the *p*-entailments: because the *p*-entailment of I is O, the same ordering of end terms as found in the *min* conclusion is predicted for both AI1 and AI3. The point we wish to emphasize is that the simplicity of the strategy means that acquiring it requires only minimal experience of reasoning syllogistically. However, this is clearly not a strategy that can be generalized from other modes of reasoning where the problem simply does not arise. Moreover, if you agree with Russell (and disagree with Johnson-Laird), then one might question whether even this minimal level of experience is available before people enter the reasoning lab.

However, we argue that many psychological processes involve implicit syllogistic reasoning which must be going on from a very early age as more knowledge about the world is slowly acquired. The purpose of acquiring world knowledge is to guide expectations. Most often, that knowledge will be acquired in different contexts and so, to derive the appropriate expectations, this information will need to be combined somehow. Combining stereotypical information is a paradigm case. As people meet other people throughout their lives, in the flesh and in fiction, they develop stereotypical information very early on. So, for example, you may come to believe that *Few skydivers are boring* but that *All accountants are boring.* If the issue of the occupations of skydivers arose, you could use this information rapidly to form the expectation that *Few skydivers are accountants.* You are likely

[16] By ‘‘similar’’ we only mean that a particular conclusion is more similar to a particular premise than to a universal premise. The only exception to this similarity based rule concerns the AE and EA syllogisms where the *p*-entailment, O, is not similar to either premise.

to perform this inference rapidly and implicitly and even be able to justify your conclusion: in general, accountants are boring and skydivers are not. However, you may only become aware that you are implicitly committed to this expectation when you are surprised on being told that the skydiver you just met was an accountant. People must be drawing implicit inferences of this form to generate their expectations about other people all the time. Rarely do these inferences receive an explicit formulation. Nonetheless, we suggest that they provide a rich backdrop of implicit experience from which people acquire their short-cut heuristics for deriving syllogistic conclusions.

Representational issues. A final possible theoretical objection to PHM is that the heuristics we have proposed do not, at least on the face of it, involve manipulating people's representations of the meanings of the premises of a syllogistic argument. This, it may be argued, contrasts with other accounts of syllogistic reasoning that account for the derivation of syllogistic conclusions by manipulating these internal representations. This fact may be felt to argue that accounts such as mental logics and mental models are, a priori, more plausible than PHM. However, as our discussion of implicit syllogistic reasoning reveals, applying these heuristics does presuppose that people represent the meanings of the premises that enter into syllogistic arguments. To identify the *min*-premise (and, therefore, the *max*-premise) requires that their probabilities, which according to our account provide the meanings of these premises, are represented in the cognitive system (although they may be represented in a more coarse-grained way than required by probability theory; see, e.g., Pearl, 1988). Moreover, the cognitive system must be able to convert these probabilities into structured representations containing the quantifier dictated by our probabilistic semantics and the arrangement of end and middle terms. Only then can the heuristics that make up PHM be applied to implicitly derive syllogistic conclusions. Consequently, PHM does require semantic representations that capture people's understanding of syllogistic premises. All that PHM denies is that people need to perform complex derivations over these representations to generate syllogistic conclusions.

Empirical Issues

We consider four empirical issues: individual differences, belief bias, multiply quantified sentences, and syllogisms using ''only.''

Individual differences. Ford (1994) argues from protocol data that some people use *verbal,* and others use *spatial,* representations of syllogistic arguments. We should expect different strategies to show up in different patterns of normal performance in syllogistic reasoning. It might, therefore, seem surprising that PHM can provide such an accurate overall data fit. One explanation is that the apparent *verbal/spatial* difference does not arise in the standard data because people are forced to change their reasoning strategy when a protocol must be generated. Thus, in terms of PHM, the difference would be located in the test procedure. An alternative possibility is that some correlate of the *verbal/spatial* difference may arise from individual differences in the generate procedures, which PHM does not currently allow. For example, there could be subtle differences in the heuristics people use, in the weightings given to *max*, or in the sophistication of the attachment heuristic and so on. Indeed, assumptions about domain size could even affect the very informational ordering that people adopt. Consequently, there seems just as much scope for explaining individual differences in PHM as in other accounts. Finally, as noted in the discussion of Polk and Newell (1995), modeling data from individual participants presents an important future direction for PHM.

Stenning and Yule (1997; see also Stenning & Oberlander, 1995) have shown how algorithms for syllogistic reasoning based on both mental logic and mental models are both instances of *individual identification algorithms*. These algorithms rely on the *case identifiability property* of syllogisms whereby a valid conclusion specifies a unique type of individual that must exist. In particular, Stenning and Yule have shown a novel way of predicting figural effects at the computational level without relying on the limitations of some putative working memory store. They identify the source premise as the source of the existential force of a valid conclusion. They suggest that the end term of the *source* premise will invariably be chosen as the subject term of the conclusion. Importantly, this hypothesis also makes predictions for figures 2 and 3. Stenning and Yule (1997) report important evidence from a novel version of the syllogistic reasoning task that is consistent with the source premise hypothesis. Although this is a very interesting development, we have some problems with these proposals. First, the concept of a source premise only applies to the valid syllogisms, ''the source premise for some valid individual conclusion is any premise which makes an existential assertion, from which the existence of that individual can be inferred,'' (see Stenning & Yule, 1997, pp. 126–127). Consequently, figural effects cannot be predicted for the invalid syllogisms. But recall that, out of 56 syllogisms for which attachment makes a prediction, 54 conformed to attachment, and yet only 27 of these are logically valid (Stenning and Yule allow existential presupposition). This leaves 27 syllogisms for which our meta-analysis reveals a figural preference that is unexplained by the ''source-founding hypothesis.'' Second, this explanation of figural effects could not account for the figural effects observed for the *p*-entailments which, when there is no response competition, are drawn as frequently as the standard inferences, as Rips's (1994) data reveal. Of course, the *p*-entailments (or, indeed, the Gricean implicatures; Grice, 1975) are not the product of deductive inferences, and so, by definition, identify the existence of individuals that can *not* be deductively inferred. Third, the concept of case identifiability, on which Stenning and Yule's (1997) analysis relies, does not apply to syllogisms containing generalized quantifiers, where, according to our probabilistic account, the existence of particular individuals is not certain, just more or less likely.

Belief bias. Belief bias occurs when the believability of the materials influence reasoning performance. Although first demonstrated in syllogistic reasoning (Morgan & Morton, 1944; Henle & Michael, 1956), belief bias effects arise in a wide range of experimental tasks (e.g., Cheng & Holyoak, 1985; Griggs & Cox, 1982; Griggs & Newstead, 1982; Wason & Brooks, 1979; Wason & Shapiro, 1971). A natural explanation is that, in everyday reasoning, people use all their knowledge to arrive at the most plausible conclusions. But in experimental reasoning tasks, participants must reason only from the given premises, ignoring background knowledge. Therefore, belief bias may simply reflect the inability of participants to disengage their everyday reasoning strategies in any reasoning context (Fodor, 1983; Oaksford & Chater, 1991; Pylyshyn, 1987) and, therefore, does not distinguish between alternative accounts of syllogistic reasoning.

Multiply quantified sentences. Johnson-Laird, Byrne, and Tabossi (1989) provide another line of related work. Consider the argument: ''None of the Princeton letters is in the same place as any of the Cambridge letters,'' ''All of the Cambridge letters are in the same place as all of the Dublin letters,'' therefore ''None of the Princeton letters is in the same place as any of the Dublin letters.'' Such reasoning appears similar to syllogistic reasoning. However, each premise and the conclusion contains (i) two quantifiers and (ii) a relation (in the example, ''is in the same place as''), which increases the logical complexity significantly. We have no general theory of the representation of relations over which we could compute informativeness. In this respect, MMT does better than PHM, because Johnson-Laird, Byrne, and Tabossi (1989) have extended mental models to handle problems such as that above, arguing that the number of mental models predicts problem difficulty. However, they deal with only one equivalence relation: ''in the same place as,''[17] rather than giving a general account of quantified relational reasoning. Further, an observation of Greene (1992) points to a possible informational account of these data. Greene (1992) observed that, as we have found for syllogisms, conclusion type reflects the distinction between one and multiple model problems—the only valid conclusion for multiple model problems was of the form ''None of the X are related to some of the Y.'' Intuitively, such conclusions seem particularly uninformative. This observation may be a starting point for an informational account, but presently these data provide most support for mental models.

[17] Equivalence relations are reflexive (xRx), transitive (xRy and yRz implies xRz), and symmetric (xRy implies yRx).

''Only'' reasoning. Johnson-Laird and Byrne (1989) have used ''Only'' (N), which is equivalent to A (i.e., Only X are Y $\Leftrightarrow$ All Y are X), in syllogistic reasoning. A critical finding is that NN syllogisms tend to lead to A conclusions, i.e., the conclusion quantifier is not of the same type as either premise, which can't happen by *min.* However, *p*-entailments do allow this to happen. They are typically less common, because *p*-entailments do not imply perfect logical entailment but only that the *p*-entailment follows with some probability (or that it is less informative than the standard conclusion and, hence, it is less likely to be picked). According to PHM, logical equivalence between different quantifiers, e.g., A and N, can be regarded as an especially strong *p*-entailment. A crucial asymmetry between A and N is that A is much more frequently used to express class inclusion. Therefore, a strong bias would be expected for people to give conclusions using the familiar A form rather than the less familiar N form. So although the *p*-entailment here is bidirectional, it will occur much more frequently in one direction than the other. This is a matter of experience with natural language usage, and not because of logical factors. This explanation gives rise to the pattern of data in Johnson-Laird and Byrne (1989).

Conclusions

Most human reasoning must deal with the uncertainty of the world in which we live. Consequently, we have argued that probability theory rather than logic provides a more appropriate computational level theory of human reasoning. In this paper, we have shown that it is possible to extend the probabilistic approach in to a core area of research in deductive reasoning, the syllogism. PHM provides a probabilistic, rather than a deductive, framework for syllogistic inference, and a set of fast and frugal heuristics which are justified by this probabilistic framework. There may be other sets of heuristics which provide a better approximation to probabilistic validity, and thus a richer model of human syllogistic reasoning. Searching alternative sets of heuristics, both for empirical adequacy and for approximation to probabilistic norms (e.g., Gigerenzer & Goldstein, 1996; McKenzie, 1994), is an important direction for future research. Currently, PHM provides more comprehensive fits to past data than alternative accounts, and is confirmed by two experiments with syllogisms involving Most and Few which are beyond the scope of other accounts. We believe that fast, frugal, and rational probabilistic heuristics will be important in explaining apparent biases in performance on laboratory reasoning tasks in other domains of reasoning. When confronted with a laboratory reasoning task, people are simply applying the strategies that they have learned to deal with their uncertain world.

APPENDIX A: DERIVATION OF THE INFORMATIVENESS ORDERING

The high prevalence of E statements is one end of a continuum. There will be a decreasing distribution (Fig. A1) of true statements between arbitrarily chosen natural language predicates as the degree of category overlap increases. In Fig. A1, the frequency of true E statements (which is over half of all statements) is represented by the filled arrow; for A statements, this frequency is represented by the small square filled box; for F and M statements, it is given by the areas marked in white. In Fig. A1, area Z does not correspond to a quantifier. The frequency with which I statements are true is represented by the union of regions F, Z, M, and A; the frequency with which O statements are true is represented by the union of regions E, F, Z, and M.

Rarity allows a complete informativeness order. In Fig. A1, A is smaller than M, and F is larger than M, because of the rapid decay in the frequencies as $P(Y|X)$ increases. Because we assume that E applies to more than half of all statements, and I includes all the other statements, I is smaller than E. This provides the following informativeness ordering: I(A) $>$ I(M) $>$ I(F) $>$ I(I) $>$ I(E) $\gg$ I(O). Notice that I(O) is far less informative than all other statements, because it is only false in the rare event that an A statement is true.

Similarity Space Model

We now justify this ordering by considering a model of the organization of natural language properties. Many models of categorization assume that

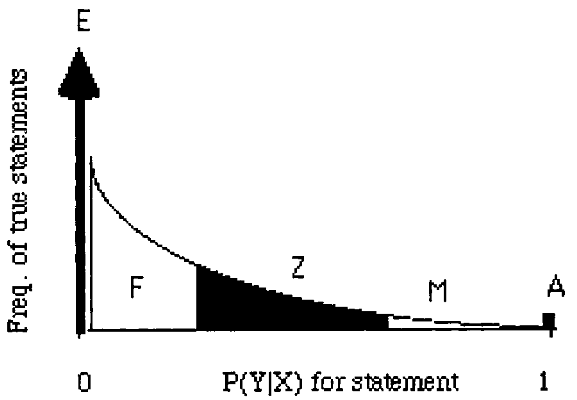

FIG. A1. Schematic diagram of frequency of true statements as a function of P(Y|X). The frequency of E statements (which is over half of all statements) is represented by the filled arrow. The frequency of A statements is represented by the small square filled box. The frequencies of the F and M statements are given by the areas marked in white. The remaining shaded area, Z, does not correspond directly to any particular quantifier.

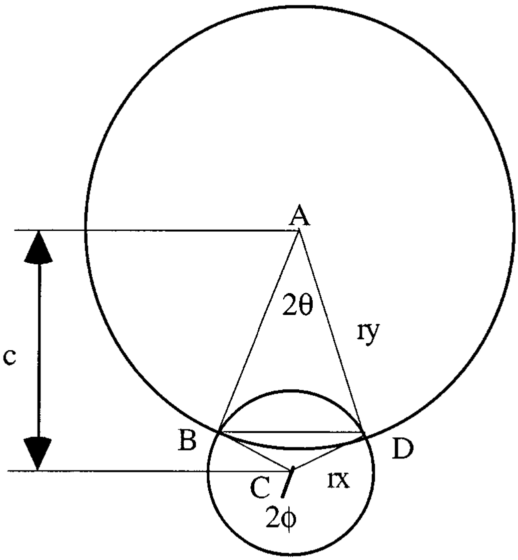

FIG. A2. Geometry of the intersection between two circles (see text for explanation).

categories are convex regions of a similarity space (e.g., Rosch, 1978; Fried & Holyoak, 1984). Here, rarity means that typical categories occupy a small convex region, so that many categories will not overlap. Hence, E-statements will frequently be true, whatever the size of the domain. Therefore, E-statements will be relatively uninformative, as in our information order.

Simplifying, we assume that categories correspond to n-dimensional spheres. n is likely to be high, but for simplicity we only consider the 2D case. Centers of categories are assigned randomly in the space with uniform probability. We ignore ''edge-effects,'' where some parts of a category might fall outside the similarity space, because we assume that the entire space is large in relation to the categories.

Without loss of generality, we assume that one of the circles, representing category X, has been placed in the similarity space (Fig. A2). The probability that the second circle, representing category Y, is placed such that a statement type Ω(X, Y) is true, is given by the ratio of the area for which it holds. The only aspect of this calculation which is complex is the derivation of the formula for the overlap between two circles, of radii r_x and r_y, whose centers are a distance c apart.

The area of intersection consists of two regions, lying on each side of the line BD. The area of the upper region can be calculated as the difference between the area of the sector BCD and the area of the triangle BCD. Similarly, the area of the lower region can be calculated as the difference between the area of the sector BAD and the area of the triangle BAD. Consequently, the area of overlap is the sum of these two differences. This can be written as $\theta r_x^2 + \phi r_y^2 - r_x^2 \sin\theta \cos\theta - r_y^2 \sin\phi \cos\phi$. To express this purely in terms of the radii r_x and r_y and the distance between the centers, c, we can exploit the following constraints: $c = r_x \cos\theta + r_y \cos\phi$ and $r_x \sin\theta = r_y \sin\phi$. Elementary algebraic manipulation leads to the following expression for the area of overlap:

$$r_x^2 \arcsin\left(\frac{r_y}{r_x}\sqrt{1 - Z^2}\right) + r_y^2 \arccos(Z) - r_y c\sqrt{1 - Z^2}, \qquad \text{(A1)}$$

where

$$Z = \frac{r_y^2 - r_x^2 + c^2}{2cr_y}.$$

Let us assume that the circle representing the predicate term, Y, is placed in the plane first. Whether the statement Ω (X, Y) is true depends on where the center of the second circle, representing the subject term X, is placed. The probability that a statement is true is proportional to the ratio of the area of this region to the total space.

There are two cases. The first, in which the predicate term Y is larger than the subject term X, is shown in Fig. A3. An A-statement, A(X, Y), is true when X is entirely included in Y, which requires that the center of X is in a small circular region around the center of Y. M(X, Y) is true when a proportion $1 - \Delta$ or greater of X overlaps Y, but X is not completely included in Y. This is true if the center of X falls within an annulus within Y, the width of which depends on Δ. F(X, Y) is true when a proportion Δ or less of X overlaps Y, when the center of X falls within an annulus enclosing Y, the width of which again depends on Δ. I(X, Y) is true if the center of X falls within a circle centered on the center of Y, whose radius is the sum of the radii of X and of Y. E(X, Y) is true when an I(X, Y) is false, which corresponds to the complement of the region where I(X, Y) is true. Similarly, O(X, Y) corresponds to the complement of the region where A(X, Y) is true. Figure A3 shows that the size of the shaded areas mirrors the standard information order, where small areas correspond to informative statements.

To quantify informativeness in this model requires estimates of the sizes of the shaded regions given values for the sizes of the circles X and Y and the value of Δ. Δ determines the proportion of X that overlaps with Y and, hence, determines values for the radii of the inner and outer circles of the annuli for quantifiers M and F. We set the radius of X to 1 and the radius

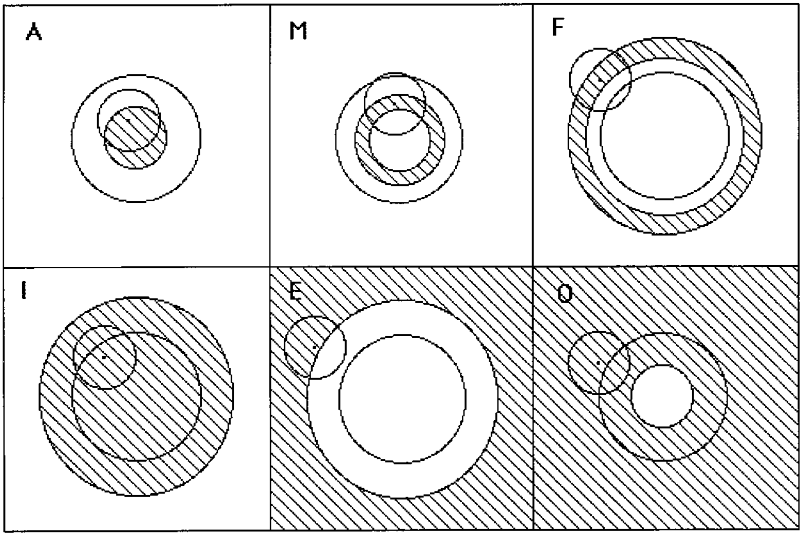

FIG. A3. The areas (shaded) in which the quantified statements $\Omega(X, Y)$ are true for randomly placed regions in similarity space, for the case in which the region for Y is larger than the region for X. Regions are assumed to be circular. The square box, which represents the rest of the whole similarity space, is not drawn to scale.

of Y to 2, and set Δ to .3, giving the standard information order: I(A) = 4.99, I(M) = 4.35, I(F) = 3.00, I(I) = 1.83, I(E) = 0.48, I(O) = 0.05. This order is extremely parameter *in*sensitive. The only departure possible is that I(M) > I(A) if the circles corresponding to X and Y are close in size.

In the second case Y is smaller than X (Fig. A4). A(X, Y) is never true because X can never be included in Y and, hence, O(X, Y) is always true. M(X, Y) is also never true because the proportion of X that overlaps Y can never be greater than $1 - \Delta$. The regions for F(X, Y), I(X, Y), and E(X, Y) are calculated as before. Quantitatively, we set the radius of X to be 2, the radius of Y to be 1, and Δ to be .3. A(X, Y) and M(X, Y) are never true and, hence, informativeness is undefined, and the other quantifiers mirror the standard order: I(F) = 1.83; I(I) = 1.83; I(E) = .48, I(O) = 0. Here, I(F) and I(I) have the same value, because even if *all* the Y are X then only .25 of the X are Y. Consequently, if X and Y overlap (i.e., I(X, Y) is true), then few of the X's are Y (i.e., F(X, Y) is true). In general, I(F) $\geq$ I(I), as in our information order.

We have justified the standard informativeness order and the noninformativeness of O. Our analysis allows us to formulate a specific version of the *min*-heuristic outlined above, that participants select the type of the least informative premise as the conclusion type, using the order A > M > F > I > E $\gg$ O.

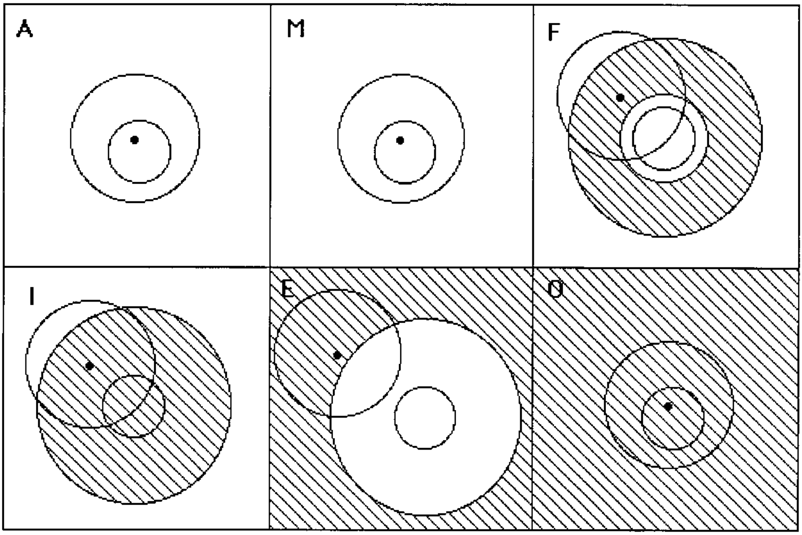

FIG. A4. The areas (shaded) in which the quantified statements Ω(X, Y) are true for randomly placed regions in similarity space, for the case in which the region for X is larger than the region for Y. Regions are assumed to be circular. The square box, which represents the rest of the whole similarity space, is not drawn to scale.

APPENDIX B: PROBABILISTICALLY VALID SYLLOGISTIC INFERENCE

This Appendix derives the notion of probabilistic validity. Syllogistic reasoning involves relating the end terms X and Z, via the middle term Y. In our probabilistic semantics, this involves using two probabilistic statements, relating X to Y and Z to Y, to derive a conclusion relating X to Z. We assume that there is no other relation between X and Z, i.e., the end terms are independent, conditional on the middle term: $\mathrm{P}(X, Y, Z) = \mathrm{P}(Y)\ \mathrm{P}(X|Y)\ \mathrm{P}(Z|Y)$ (where X, Y and Z are ''literals'' in the standard logical sense, i.e., they stand for themselves or their negations).

The four figures correspond to different relations between premise terms, shown formally using dependency diagrams in Fig. B1 (Pearl, 1988). X, Y and Z are nodes; dependencies are arrows between nodes indicating the appropriate conditional probabilities (i.e., $P(Y|X)$ corresponds to $X \rightarrow Y$). X and Z are not directly connected, representing conditional independence. Nodes receiving no arrow represent unconditional probabilities (e.g., $P(X)$). Five probabilities parameterize each diagram. Two parameters are set by the premise quantifiers, e.g., AA4: *All X are Y, All Y are Z* specifies: $P(Y|X) = 1$, $P(Z|Y) = 1$. $P(X)$, $P(Y|\overline{X})$, $P(Z|\overline{Y})$ vary freely between 0 and 1, but may be constrained indirectly by the premises. For example, *Some X are not-Y*

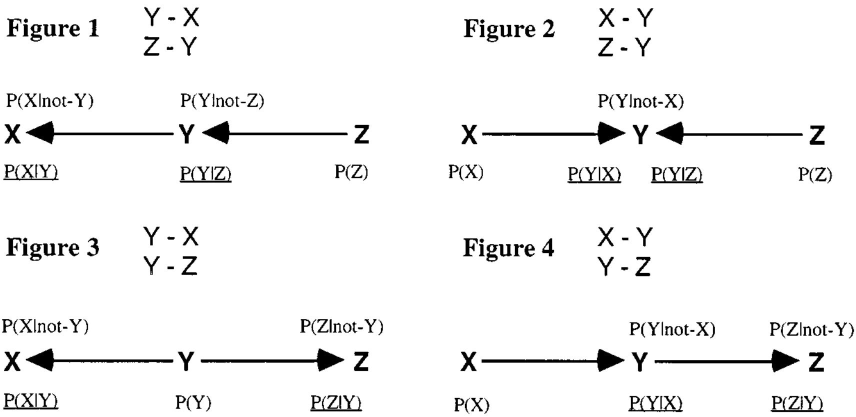

FIG. B1. Probability models for the four syllogistic figures and their parameters. For each syllogistic figure, arrows indicate dependencies between terms. The direction of the arrow corresponds to the appropriate conditional probability (i.e., P(Y|X) corresponds to X → Y). The nodes corresponding to the end terms are not directly connected, which means they are conditionally independent, given the middle term. The probabilities shown are the parameters for each model. Nodes which do not receive an arrow from other nodes are associated with unconditional probabilities (e.g., $P(X)$).

is represented as $P(Y|X) < 1$. But it also implies that there are some *X*, ($P(X) > 0$), and that not everything is *Y*, ($P(Y) < 1$). These constraints may affect *p*-validity.

A final technicality concerns I(X, Y) and E(X, Y), which we represent as joint rather than as conditional probabilities. We take *Some X are Y* to mean $P(X, Y) > 0$, and *No X are Y* to mean $P(X, Y) = 0$. *Some X are Y* means that $P(X|Y) > 0$, which implies the additional constraints $P(X) > 0$ and $P(Y) > 0$, whereas the single statement $P(X, Y) > 0$ captures them both. *No X are Y* means that $P(X, Y) = 0$ and $P(Y|X) = 0$ are equivalent, except when $P(X) = 0$, where only the joint probability is undefined. We avoid adding this constraint by using joint probabilities. Using joint probabilities is for convenience of calculation only.

To determine whether a conclusion is *p*-valid, we derive expressions for $P(X, Z)$ (to find I- and E-conclusions), and $P(X|Z)$ and $P(Z|X)$ (to find A-, M-, F-, and O-conclusions) using the parameters for the different figures, and determine whether there are constraints on the relation between *X* and *Z* for all parameter values. If $P(X, Z) > 0$, then an I-conclusion follows. If $P(X, Z) = 0$, then an E-conclusion follows. If $P(X|Z) = 1$, then A follows (*All Z are X*); alternatively, if $P(Z|X)$ is 1, then a different A follows (*All X are Z*). If $1 - \Delta \leq P(X|Z) < 1$, then M follows (*Most Z are X*); alternatively, if $1 - \Delta \leq P(Z|X) < 1$, then a different M follows (*Most X are Z*). If $0 < P(X|Z) \leq \Delta$, then an F follows (*Few Z are X*); alternatively, if $0 < P(X|Z) \leq \Delta$, then a different F follows (*Few X are Z*). If $P(X|Z) < 1$, then an

O follows (*Some Z are not X*); alternatively, if $P(X|Z) < 1$, then a different O follows (*Some X are not Z*).

As an example, consider the syllogism, AI1: *All Y are X, Some Z are Y.* These premises translate into the following constraints on the probabilistic parameters for figure 1: $P(X|Y) = 1$, $P(Y, Z) > 0$. The remaining parameters are $P(Z)$, $P(Y|\bar{Z})$, $P(X|\bar{Y})$. Notice that $P(Z) \geq P(Y, Z) > 0$, so that $P(Z) > 0$. We want to express the three probabilities relevant to whether a conclusion follows ($P(X, Z)$, $P(X|Z)$, and $P(Z|X)$), in terms of the parameters for figure 1. So, using these parameters, shown in Fig. B1, we derive the expressions for each of the conclusion probabilities:

$$P(X, Z) = P(X|Y)P(Y, Z) + P(X|\bar{Y})(P(Z) - P(Y, Z)) > 0 \qquad \text{I} \tag{B1}$$

$$P(X, Z) = \frac{1}{P(Z)}(P(X|Y)P(Y, Z) + P(X|\bar{Y})(P(Z) - P(Y, Z))) > 0 \qquad \text{NVC} \tag{B2}$$

$$P(Z, X) = \frac{P(X|Y)P(Y, Z) + P(X|\bar{Y})(P(Z) - P(Y, Z))}{\begin{pmatrix} P(X|Y)P(Y, Z) + P(X|\bar{Y})(P(Z) - P(Y, Z)) \\ + P(X|Y)P(Y|\bar{Z})(1 - P(Z)) \\ + P(X|\bar{Y})((1 - P(Z))(1 - P(Y|\bar{Z}))) \end{pmatrix}} > 0 \qquad \text{NVC} \tag{B3}$$

Substituting in the constraints derived from the premises, it follows that all these probabilities are greater than 0. We show the resulting conclusion (A, M, F, I, E, or O) or NVC if no conclusion follows by each expression. $P(X, Z)$ must be greater than 0 and, therefore, the I-conclusions *Some X are Z* and *Some Z are X* follow *p*-validly. The only constraints on the conditional probabilities is that they are both greater than 0, but no further conclusions follow from this. We have performed similar derivations for all 144 possible syllogisms made up of the 36 possible pairs of premises and the four figures.

APPENDIX C: META-ANALYSIS OF SYLLOGISTIC REASONING EXPERIMENTS

TABLE C1
Meta-Analysis Showing the Percentage of Times Each Conclusion Was Drawn in Each of the Five Studies D1, D2, J-LS1, J-LS2, and J-LB Weighted by Sample Size, and the Predictions of PHM

	Conclusion type								
	A		I		E		O		NVC
Syllogism	Data	Model	Data	Model	Data	Model	Data	Model	Data
AA1	***89.87***	***70.14***	*5.06*	*10.76*	0.00	1.22	0.00	1.22	1.90
AA2	***58.23***	***70.14***	*8.23*	*10.76*	1.27	1.22	1.27	1.22	28.48
AA3	***56.96***	***70.14***	*29.11*	*10.76*	0.00	1.22	0.00	1.22	13.92
AA4	***75.32***	***70.14***	*16.46*	*10.76*	1.27	1.22	0.63	1.22	4.43
AI1	0.00	1.22	***92.41***	***70.14***	2.53	1.22	*2.53*	*10.76*	2.53
AI2	0.00	1.22	***56.96***	***70.14***	2.53	1.22	*10.76*	*10.76*	29.11
AI3	1.27	1.22	***88.61***	***70.14***	1.27	1.22	*2.53*	*10.76*	4.43
AI4	0.00	1.22	***70.89***	***70.14***	0.00	1.22	*1.27*	*10.76*	26.58
IA1	0.00	1.22	***71.52***	***70.14***	0.00	1.22	*6.33*	*10.76*	21.52
IA2	13.29	1.22	***48.73***	***70.14***	2.53	1.22	*12.03*	*10.76*	31.01
IA3	1.90	1.22	***84.81***	***70.14***	0.63	1.22	*3.80*	*10.76*	5.70
IA4	0.00	1.22	***91.14***	***70.14***	1.27	1.22	*0.63*	*10.76*	4.43
AE1	0.00	1.22	2.53	1.22	***59.49***	***70.14***	*5.70*	*10.76*	31.65
AE2	0.00	1.22	0.00	1.22	***87.97***	***70.14***	*1.27*	*10.76*	5.70
AE3	0.00	1.22	1.27	1.22	***61.39***	***70.14***	*13.29*	*10.76*	22.78
AE4	0.00	1.22	2.53	1.22	***86.71***	***70.14***	*1.90*	*10.76*	2.53
EA1	0.00	1.22	1.27	1.22	***86.71***	***70.14***	*2.53*	*10.76*	4.43
EA2	0.00	1.22	0.00	1.22	***88.61***	***70.14***	*3.16*	*10.76*	3.80
EA3	0.00	1.22	0.00	1.22	***63.92***	***70.14***	*21.52*	*10.76*	12.03
EA4	1.27	1.22	2.53	1.22	***61.39***	***70.14***	*8.23*	*10.76*	24.05
AO1	1.27	1.22	*6.33*	*10.76*	*1.27*	*1.22*	***56.96***	***70.14***	31.01
AO2	0.00	1.22	*6.33*	*10.76*	2.53	1.22	***67.09***	***70.14***	21.52
AO3	0.00	1.22	*10.13*	*10.76*	0.00	1.22	***66.46***	***70.14***	18.99
AO4	0.00	1.22	*5.06*	*10.76*	2.53	1.22	***71.52***	***70.14***	19.62
OA1	0.00	1.22	*3.16*	*10.76*	2.53	1.22	***67.72***	***70.14***	23.42
OA2	0.00	1.22	*11.39*	*10.76*	5.06	1.22	***56.33***	***70.14***	22.78
OA3	0.00	1.22	*14.56*	*10.76*	3.16	1.22	***68.99***	***70.14***	8.23
OA4	1.27	1.22	*2.53*	*10.76*	6.33	1.22	***27.22***	***70.14***	32.91
II1	0.00	1.22	***40.51***	***31.11***	2.53	1.22	*4.43*	*10.76*	50.63
II2	1.27	1.22	***42.41***	***31.11***	3.16	1.22	*2.53*	*10.76*	49.37
II3	0.00	1.22	***24.05***	***31.11***	2.53	1.22	*1.27*	*10.76*	71.52
II4	0.00	1.22	***42.41***	***31.11***	0.00	1.22	*1.27*	*10.76*	56.96
IE1	1.27	1.22	1.27	1.22	***22.15***	***31.11***	*16.46*	*10.76*	56.96
IE2	0.00	1.22	0.00	1.22	***39.24***	***31.11***	*30.38*	*10.76*	27.22
IE3	0.00	1.22	1.27	1.22	***29.75***	***31.11***	*32.91*	*10.76*	34.81
IE4	0.00	1.22	0.63	1.22	***27.85***	***31.11***	*44.30*	*10.76*	24.05
EI1	0.00	1.22	5.06	1.22	***14.56***	***31.11***	*66.46*	*10.76*	13.92
EI2	1.27	1.22	1.27	1.22	***20.89***	***31.11***	*51.90*	*10.76*	21.52
EI3	0.00	1.22	6.33	1.22	***15.19***	***31.11***	*48.10*	*10.76*	27.22

TABLE C1—*Continued*

	Conclusion type								
	A		I		E		O		NVC
Syllogism	Data	Model	Data	Model	Data	Model	Data	Model	Data
EI4	0.00	1.22	1.90	1.22	***31.65***	***31.11***	*26.58*	*10.76*	37.34
IO1	2.53	1.22	*3.80*	*10.76*	1.27	1.22	***30.38***	***31.11***	60.76
IO2	1.27	1.22	*5.06*	*10.76*	3.80	1.22	***36.71***	***31.11***	48.73
IO3	0.00	1.22	*9.49*	*10.76*	1.27	1.22	***29.11***	***31.11***	58.86
IO4	0.00	1.22	*5.06*	*10.76*	1.27	1.22	***44.30***	***31.11***	46.84
OI1	3.80	1.22	*6.33*	*10.76*	0.00	1.22	***35.44***	***31.11***	53.80
OI2	0.00	1.22	*8.23*	*10.76*	2.53	1.22	***35.44***	***31.11***	50.00
OI3	1.27	1.22	*8.86*	*10.76*	1.27	1.22	***31.01***	***31.11***	56.33
OI4	2.53	1.22	*7.59*	*10.76*	1.90	1.22	***29.11***	***31.11***	55.06
EE1	0.00	1.22	1.27	1.22	***34.18***	***18.78***	*1.27*	*10.76*	60.76
EE2	2.53	1.22	2.53	1.22	***13.92***	***18.78***	*3.16*	*10.76*	77.22
EE3	0.00	1.22	0.00	1.22	***17.72***	***18.78***	*2.53*	*10.76*	77.85
EE4	0.00	1.22	2.53	1.22	***31.01***	***18.78***	*0.63*	*10.76*	63.29
EO1	1.27	1.22	*7.59*	*10.76*	7.59	1.22	***23.42***	***18.78***	59.49
EO2	0.00	1.22	*13.29*	*10.76*	6.96	1.22	***11.39***	***18.78***	65.19
EO3	0.00	1.22	*0.00*	*10.76*	8.86	1.22	***28.48***	***18.78***	58.23
EO4	0.00	1.22	*5.06*	*10.76*	8.23	1.22	***12.03***	***18.78***	70.25
OE1	1.27	1.22	*0.00*	*10.76*	13.92	1.22	***5.06***	***18.78***	77.22
OE2	0.00	1.22	*7.59*	*10.76*	10.76	1.22	***16.46***	***18.78***	62.66
OE3	0.00	1.22	*5.06*	*10.76*	12.03	1.22	***17.72***	***18.78***	60.76
OE4	0.00	1.22	*18.99*	*10.76*	9.49	1.22	***13.92***	***18.78***	56.33
OO1	1.27	1.22	*7.59*	*10.76*	1.27	1.22	***22.15***	***18.04***	66.46
OO2	0.00	1.22	*16.46*	*10.76*	5.06	1.22	***10.13***	***18.04***	68.35
OO3	1.27	1.22	*6.33*	*10.76*	0.00	1.22	***15.19***	***18.04***	76.58
OO4	1.27	1.22	*4.43*	*10.76*	0.63	1.22	***24.68***	***18.04***	64.56

Note. The label for the logically valid syllogisms is in italics. The numbers that are in italics are those predicted by the *min*-heuristic with *p*-entailments, as outlined in the text. The numbers which are in bold are those which would also be predicted *without p*-entailments. Where all responses for a syllogism do not sum to a 100, this is because, in some studies, other irrelevant responses were made.

APPENDIX D: EXPERIMENT 1 RESULTS

TABLE D1
The Frequency (%) of Each Conclusion Type for Each Syllogism in Experiment 1 Using the Quantifiers AMFO

	Conclusion type								
	A		M		F		O		NVC
Syllogism	Data	Model	Data	Model	Data	Model	Data	Model	Data
AA1	***85.00***	***67.03***	0.00	5.27	0.00	5.27	10.00	5.27	5.00
AA2	***60.00***	***67.03***	5.00	5.27	10.00	5.27	20.00	5.27	20.00
AA3	***65.00***	***67.03***	0.00	5.27	5.00	5.27	15.00	5.27	20.00
AA4	***55.00***	***67.03***	15.00	5.27	10.00	5.27	15.00	5.27	25.00
AM1	5.00	5.27	***85.00***	***67.03***	0.00	5.27	*40.00*	*38.93*	0.00
AM2	5.00	5.27	***60.00***	***67.03***	5.00	5.27	*35.00*	*38.93*	20.00
AM3	10.00	5.27	***60.00***	***67.03***	10.00	5.27	*40.00*	*38.93*	10.00
AM4	5.00	5.27	***70.00***	***67.03***	10.00	5.27	*35.00*	*38.93*	5.00
MA1	5.00	5.27	***65.00***	***67.03***	10.00	5.27	*40.00*	*38.93*	5.00
MA2	10.00	5.27	***50.00***	***67.03***	10.00	5.27	*35.00*	*38.93*	15.00
MA3	5.00	5.27	***65.00***	***67.03***	15.00	5.27	*50.00*	*38.93*	5.00
MA4	20.00	5.27	***55.00***	***67.03***	15.00	5.27	*30.00*	*38.93*	5.00
AF1	5.00	5.27	0.00	5.27	***85.00***	***67.03***	*40.00*	*38.93*	0.00
AF2	5.00	5.27	0.00	5.27	***85.00***	***67.03***	*25.00*	*38.93*	10.00
AF3	5.00	5.27	0.00	5.27	***80.00***	***67.03***	*35.00*	*38.93*	5.00
AF4	5.00	5.27	0.00	5.27	***75.00***	***67.03***	*25.00*	*38.93*	15.00
FA1	10.00	5.27	0.00	5.27	***80.00***	***67.03***	*30.00*	*38.93*	5.00
FA2	5.00	5.27	5.00	5.27	***70.00***	***67.03***	*30.00*	*38.93*	15.00
FA3	10.00	5.27	5.00	5.27	***60.00***	***67.03***	*45.00*	*38.93*	0.00
FA4	5.00	5.27	5.00	5.27	***65.00***	***67.03***	*45.00*	*38.93*	5.00
AO1	5.00	5.27	*25.00*	*17.01*	*10.00*	*17.01*	***90.00***	***67.03***	5.00
AO2	5.00	5.27	*15.00*	*17.01*	*10.00*	*17.01*	***75.00***	***67.03***	10.00
AO3	5.00	5.27	*10.00*	*17.01*	*10.00*	*17.01*	***80.00***	***67.03***	15.00
AO4	5.00	5.27	*10.00*	*17.01*	*10.00*	*17.01*	***80.00***	***67.03***	15.00
OA1	10.00	5.27	*20.00*	*17.01*	*10.00*	*17.01*	***85.00***	***67.03***	10.00
OA2	5.00	5.27	*15.00*	*17.01*	*20.00*	*17.01*	***65.00***	***67.03***	10.00
OA3	5.00	5.27	*10.00*	*17.01*	*15.00*	*17.01*	***85.00***	***67.03***	10.00
OA4	10.00	5.27	*25.00*	*17.01*	*30.00*	*17.01*	***55.00***	***67.03***	10.00
MM1	0.00	5.27	***65.00***	***67.03***	15.00	5.27	*35.00*	*38.93*	15.00
MM2	0.00	5.27	***45.00***	***67.03***	20.00	5.27	*30.00*	*38.93*	25.00
MM3	0.00	5.27	***50.00***	***67.03***	5.00	5.27	*50.00*	*38.93*	10.00
MM4	0.00	5.27	***50.00***	***67.03***	15.00	5.27	*35.00*	*38.93*	15.00
MF1	0.00	5.27	0.00	5.27	***75.00***	***67.03***	*25.00*	*38.93*	10.00
MF2	0.00	5.27	10.00	5.27	***60.00***	***67.03***	*35.00*	*38.93*	15.00
MF3	0.00	5.27	10.00	5.27	***55.00***	***67.03***	*50.00*	*38.93*	10.00
MF4	0.00	5.27	5.00	5.27	***70.00***	***67.03***	*30.00*	*38.93*	10.00
FM1	0.00	5.27	10.00	5.27	***60.00***	***67.03***	*50.00*	*38.93*	5.00
FM2	0.00	5.27	20.00	5.27	***50.00***	***67.03***	*40.00*	*38.93*	10.00
FM3	0.00	5.27	10.00	5.27	***70.00***	***67.03***	*30.00*	*38.93*	10.00
FM4	0.00	5.27	15.00	5.27	***65.00***	***67.03***	*30.00*	*38.93*	10.00
MO1	0.00	5.27	*10.00*	*17.01*	*25.00*	*17.01*	***80.00***	***67.03***	10.00

TABLE D1—*Continued*

	Conclusion type								
	A		M		F		O		NVC
Syllogism	Data	Model	Data	Model	Data	Model	Data	Model	Data
MO2	0.00	5.27	*10.00*	*17.01*	*20.00*	*17.01*	***75.00***	***67.03***	15.00
MO3	0.00	5.27	*30.00*	*17.01*	*15.00*	*17.01*	***65.00***	***67.03***	15.00
MO4	0.00	5.27	*15.00*	*17.01*	*20.00*	*17.01*	***75.00***	***67.03***	15.00
OM1	0.00	5.27	*20.00*	*17.01*	*20.00*	*17.01*	***60.00***	***67.03***	20.00
OM2	0.00	5.27	*20.00*	*17.01*	*25.00*	*17.01*	***70.00***	***67.03***	15.00
OM3	0.00	5.27	*25.00*	*17.01*	*30.00*	*17.01*	***75.00***	***67.03***	10.00
OM4	0.00	5.27	*25.00*	*17.01*	*15.00*	*17.01*	***70.00***	***67.03***	15.00
FF1	0.00	5.27	5.00	5.27	***55.00***	***67.03***	*35.00*	*38.93*	20.00
FF2	0.00	5.27	10.00	5.27	***55.00***	***67.03***	*30.00*	*38.93*	20.00
FF3	0.00	5.27	10.00	5.27	***50.00***	***67.03***	*35.00*	*38.93*	20.00
FF4	0.00	5.27	5.00	5.27	***60.00***	***67.03***	*30.00*	*38.93*	20.00
FO1	0.00	5.27	*5.00*	*17.01*	*40.00*	*17.01*	***60.00***	***67.03***	20.00
FO2	0.00	5.27	*0.00*	*17.01*	*55.00*	*17.01*	***65.00***	***67.03***	10.00
FO3	0.00	5.27	*5.00*	*17.01*	*60.00*	*17.01*	***60.00***	***67.03***	10.00
FO4	0.00	5.27	*0.00*	*17.01*	*55.00*	*17.01*	***60.00***	***67.03***	10.00
OF1	0.00	5.27	*5.00*	*17.01*	*45.00*	*17.01*	***40.00***	***67.03***	15.00
OF2	0.00	5.27	*0.00*	*17.01*	*45.00*	*17.01*	***55.00***	***67.03***	15.00
OF3	0.00	5.27	*0.00*	*17.01*	*50.00*	*17.01*	***55.00***	***67.03***	15.00
OF4	0.00	5.27	*5.00*	*17.01*	*65.00*	*17.01*	***40.00***	***67.03***	15.00
OO1	0.00	5.27	*10.00*	*17.01*	*10.00*	*17.01*	***80.00***	***67.03***	10.00
OO2	0.00	5.27	*10.00*	*17.01*	*20.00*	*17.01*	***70.00***	***67.03***	15.00
OO3	0.00	5.27	*15.00*	*17.01*	*15.00*	*17.01*	***60.00***	***67.03***	25.00
OO4	0.00	5.27	*5.00*	*17.01*	*20.00*	*17.01*	***70.00***	***67.03***	20.00

Note. The predictions of PHM are also shown. The label for the *p*-valid syllogisms is in italics; when the *p*-valid conclusion for a syllogism was available in Experiment 1 its label is also in bold. The numbers that are in italics are those predicted by the *min*-heuristic with *p*-entailments, as outlined in the text. The numbers which are in bold are those which would also be predicted *without* *p*-entailments. Rows do not sum to 100 because more than one response was allowed.

APPENDIX E: EXPERIMENT 2 RESULTS

TABLE E1
The Frequency (%) of Each Conclusion Type for Each Syllogism in Experiment 2 using the Quantifiers MFIE

	Conclusion type								
	M		F		I		E		NVC
Syllogism	Data	Model	Data	Model	Data	Model	Data	Model	Data
MM1	***45.00***	***37.03***	10.00	12.28	*20.00*	*16.00*	15.00	12.28	20.00
MM2	***50.00***	***37.03***	15.00	12.28	*20.00*	*16.00*	5.00	12.28	25.00
MM3	***35.00***	***37.03***	10.00	12.28	*45.00*	*16.00*	5.00	12.28	15.00
MM4	***50.00***	***37.03***	5.00	12.28	*30.00*	*16.00*	5.00	12.28	20.00
MF1	15.00	12.28	***45.00***	***37.03***	*10.00*	*16.00*	10.00	12.28	30.00
MF2	20.00	12.28	***35.00***	***37.03***	*20.00*	*16.00*	10.00	12.28	25.00
MF3	15.00	12.28	***30.00***	***37.03***	*20.00*	*16.00*	10.00	12.28	30.00
MF4	15.00	12.28	***45.00***	***37.03***	*15.00*	*16.00*	5.00	12.28	35.00
FM1	0.00	12.28	***70.00***	***37.03***	*10.00*	*16.00*	10.00	12.28	20.00
FM2	0.00	12.28	***75.00***	***37.03***	*5.00*	*16.00*	15.00	12.28	20.00
FM3	0.00	12.28	***45.00***	***37.03***	*20.00*	*16.00*	15.00	12.28	30.00
FM4	0.00	12.28	***65.00***	***37.03***	*5.00*	*16.00*	15.00	12.28	25.00
MI1	*10.00*	*14.13*	*0.00*	*14.13*	***50.00***	***37.03***	15.00	12.28	30.00
MI2	*20.00*	*14.13*	*15.00*	*14.13*	***40.00***	***37.03***	5.00	12.28	30.00
MI3	*10.00*	*14.13*	*10.00*	*14.13*	***45.00***	***37.03***	10.00	12.28	30.00
MI5	*15.00*	*14.13*	*10.00*	*14.13*	***45.00***	***37.03***	10.00	12.28	30.00
IM1	*10.00*	*14.13*	*25.00*	*14.13*	***30.00***	***37.03***	10.00	12.28	30.00
IM2	*0.00*	*14.13*	*20.00*	*14.13*	***45.00***	***37.03***	5.00	12.28	30.00
IM3	*5.00*	*14.13*	*15.00*	*14.13*	***50.00***	***37.03***	5.00	12.28	35.00
IM4	*0.00*	*14.13*	*25.00*	*14.13*	***45.00***	***37.03***	10.00	12.28	30.00
ME1	15.00	12.28	10.00	12.28	10.00	12.28	***35.00***	***37.03***	40.00
ME2	5.00	12.28	35.00	12.28	10.00	12.28	***25.00***	***37.03***	35.00
ME3	10.00	12.28	20.00	12.28	15.00	12.28	***30.00***	***37.03***	35.00
ME3	15.00	12.28	15.00	12.28	5.00	12.28	***35.00***	***37.03***	40.00
EM1	5.00	12.28	35.00	12.28	20.00	12.28	***35.00***	***37.03***	15.00
EM2	5.00	12.28	50.00	12.28	5.00	12.28	***30.00***	***37.03***	20.00
EM3	5.00	12.28	20.00	12.28	15.00	12.28	***35.00***	***37.03***	35.00
EM4	5.00	12.28	35.00	12.28	10.00	12.28	***25.00***	***37.03***	35.00
FF1	5.00	12.28	***45.00***	***37.03***	*10.00*	*16.00*	10.00	12.28	30.00
FF2	0.00	12.28	***35.00***	***37.03***	*20.00*	*16.00*	25.00	12.28	35.00
FF3	10.00	12.28	***35.00***	***37.03***	*10.00*	*16.00*	20.00	12.28	35.00
FF4	0.00	12.28	***55.00***	***37.03***	*10.00*	*16.00*	15.00	12.28	30.00
FI1	*0.00*	*14.13*	*30.00*	*14.13*	***35.00***	***37.03***	15.00	12.28	30.00
FI2	*5.00*	*14.13*	*40.00*	*14.13*	***15.00***	***37.03***	20.00	12.28	35.00
FI3	*5.00*	*14.13*	*35.00*	*14.13*	***25.00***	***37.03***	20.00	12.28	30.00
FI4	*0.00*	*14.13*	*55.00*	*14.13*	***15.00***	***37.03***	5.00	12.28	35.00
IF1	*15.00*	*14.13*	*35.00*	*14.13*	***25.00***	***37.03***	10.00	12.28	30.00
IF2	*5.00*	*14.13*	*40.00*	*14.13*	***25.00***	***37.03***	20.00	12.28	30.00
IF3	*0.00*	*14.13*	*40.00*	*14.13*	***30.00***	***37.03***	5.00	12.28	35.00
IF4	*0.00*	*14.13*	*25.00*	*14.13*	***45.00***	***37.03***	5.00	12.28	30.00
FE1	0.00	12.28	20.00	12.28	15.00	12.28	***30.00***	***37.03***	45.00

TABLE E1—*Continued*

	Conclusion type								
	M		F		I		E		NVC
Syllogism	Data	Model	Data	Model	Data	Model	Data	Model	Data
FE2	5.00	12.28	15.00	12.28	10.00	12.28	***40.00***	***37.03***	40.00
FE3	5.00	12.28	10.00	12.28	15.00	12.28	***30.00***	***37.03***	45.00
FE4	5.00	12.28	10.00	12.28	15.00	12.28	***25.00***	***37.03***	50.00
EF1	15.00	12.28	5.00	12.28	15.00	12.28	***25.00***	***37.03***	45.00
EF2	10.00	12.28	15.00	12.28	25.00	12.28	***25.00***	***37.03***	35.00
EF3	5.00	12.28	10.00	12.28	20.00	12.28	***35.00***	***37.03***	35.00
EF4	10.00	12.28	15.00	12.28	15.00	12.28	***30.00***	***37.03***	35.00
II1	*0.00*	*14.13*	*10.00*	*14.13*	**50.00**	***37.03***	5.00	12.28	35.00
II2	*5.00*	*14.13*	*5.00*	*14.13*	**55.00**	***37.03***	5.00	12.28	30.00
II3	*5.00*	*14.13*	*15.00*	*14.13*	**45.00**	***37.03***	10.00	12.28	35.00
II4	*0.00*	*14.13*	*5.00*	*14.13*	**60.00**	***37.03***	5.00	12.28	30.00
IE1	10.00	12.28	0.00	12.28	20.00	12.28	***35.00***	***37.03***	40.00
IE2	5.00	12.28	15.00	12.28	25.00	12.28	***25.00***	***37.03***	45.00
IE3	0.00	12.28	10.00	12.28	35.00	12.28	***20.00***	***37.03***	40.00
IE4	5.00	12.28	20.00	12.28	20.00	12.28	***30.00***	***37.03***	35.00
EI1	0.00	12.28	20.00	12.28	30.00	12.28	***25.00***	***37.03***	30.00
EI2	5.00	12.28	5.00	12.28	35.00	12.28	***30.00***	***37.03***	35.00
EI3	0.00	12.28	10.00	12.28	30.00	12.28	***25.00***	***37.03***	40.00
EI4	5.00	12.28	30.00	12.28	20.00	12.28	***30.00***	***37.03***	25.00
EE1	15.00	12.28	5.00	12.28	20.00	12.28	***30.00***	***37.03***	35.00
EE2	15.00	12.28	10.00	12.28	20.00	12.28	***30.00***	***37.03***	40.00
EE3	10.00	12.28	0.00	12.28	20.00	12.28	***30.00***	***37.03***	45.00
EE4	15.00	12.28	0.00	12.28	20.00	12.28	***35.00***	***37.03***	40.00

Note. The predictions of PHM are also shown. The label for the *p*-valid syllogisms is in italics; when the *p*-valid conclusion for a syllogism was available in Experiment 2, its label is also in bold. The numbers that are in italics are those predicted by the *min*-heuristic with *p*-entailments, as outlined in the text. The numbers which are in bold are those which would also be predicted *without p*-entailments. Rows do not sum to 100 because more than one response was allowed.

REFERENCES

Adams, E. (1966). Probability and the logic of conditionals. In J. Hintikka & P. Suppes (Eds.), *Aspects of inductive logic.* Amsterdam, Netherlands: North Holland.

Adams, E. (1975). *The logic of conditionals: An application of probability theory to deductive logic.* Dordrecht, Netherlands: Reidel.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991a). Is human cognition adaptive? *Behavioral and Brain Sciences,* **14, 471–517.**

Anderson, J. R. (1991b). The adaptive nature of human categorisation. *Psychological Review,* **98,** 409–429.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review,* **96,** 703–719.

Bar-Hillel, Y., & Carnap, R. (1964). An outline of a theory of semantic information. In Y. Bar-Hillel (Ed.), *Language and information.* Reading, MA: Addison-Wesley.

Baron, J. (1994). *Thinking and deciding.* Cambridge: Cambridge Univ. Press.

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

Barwise, J., & Perry, J. (1983). *Situation and attitudes.* Cambridge, MA: MIT Press.

Begg, I., & Denny, J. P. (1969). Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning errors. *Journal of Experimental Psychology,* **81,** 351–354.

Begg, I., & Harris, G. (1982). On the interpretation of syllogisms. *Journal of Verbal Learning and Verbal Behaviour,* **21,** 595–620.

Belsey, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity.* New York: John Wiley and Sons.

Boole, G. (1854). *An investigation into the laws of thought.* London, UK: Macmillan.

Braine, M. D. S. (1978). On the relationship between the natural logic of reasoning and standard logic. *Psychological Review,* **85,** 1–21.

Ceraso, J., & Provitera, A. (1971). Sources of error in syllogistic reasoning. *Cognitive Psychology,* **2,** 400–410.

Chapman, L. J., & Chapman, J. P. (1959). Atmosphere effect re-visited. *Journal of Experimental Psychology,* **58,** 220–226.

Chater, N. (1993). Mental models and non-monotonic reasoning. *Behavioral and Brain Sciences,* **16,** 340–341.

Chater, N., & Oaksford, M. (1990). Autonomy, implementation and cognitive architecture: A reply to Fodor and Pylyshyn. *Cognition,* **34,** 93–107.

Chater, N., & Oaksford, M. (1993). Logicism, mental models and everyday reasoning. *Mind and Language,* **8,** 72–89.

Chater, N., & Oaksford, M. (1996). Deontic reasoning, modules and innateness: A second look. *Mind and Language,* **11,** 191–202.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology,* **17,** 391–416.

Cherniak, C. (1986). *Minimal rationality.* Cambridge, MA: MIT Press.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioural & Brain Sciences,* **4,** 317–370.

Curry, H. B. (1956). *An introduction to mathematical logic.* Amsterdam, Netherlands: Van Nostrand.

Curry, H. B., & Feys, R. (Eds.) (1958). *Combinatory logic.* Amsterdam, Netherlands: North-Holland.

de Long, H. (1971). *A profile of mathematical logic.* Reading, MA: Addison-Wesley.

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

Eells, E., & Skyrms, B. (Eds.) (1994). *Probability and conditionals: Belief revision and rational decision.* Cambridge: Cambridge Univ. Press.

Erickson, J. R. (1974). A set analysis of behavior in formal syllogistic reasoning tasks. In R. L. Solso (Ed.), *Theories of cognitive psychology: The Loyola Symposium.* Hillsdale, NJ: Erlbaum.

Evans, J. St.B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning.* Hillsdale, NJ: Erlbaum.

Evans, J. St.B. T., & Over, D. E. (1996a). Rationality in the selection task: Epistemic utility versus uncertainty reduction. *Psychological Review,* **103,** 356–363.

Evans, J. St.B. T., & Over, D. E. (1996b). *Rationality and reasoning.* Hove, East Sussex, UK: Psychology Press.

Fodor, J. A. (1983). *Modularity of mind.* Cambridge, MA: MIT Press.

Ford, M. (1994). Two modes of mental representation and problem solving in syllogistic reasoning. *Cognition,* **54,** 1–71.

Frege, G. (1879). *Begriffsschrift.* Halle, Germany: Nebert.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory and Cognition,* **10,** 234–257.

Galotti, K. M., Baron, J., & Sabini, J. P. (1986). Individual differences in syllogistic reasoning: Deduction rules or mental models. *Journal of Experimental Psychology: General,* **115,** 16–25.

Gentzen, G. (1934). Untersuchungen uber das logishce Schliessen. *Mathematische Zeitschrifft,* **39,** 176–210.

George, C. (1997). Reasoning from uncertain premises. *Thinking and Reasoning,* **3,** 161–189.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review,* **103,** 650–669.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilisitic mental models: A Brunswickian theory of confidence. *Psychological Review,* **98,** 506–528.

Glass, G., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Gordon, J., & Shortliffe, E. H. (1985). A method for managing evidential reasoning in a hierarchical hypothesis space. *Artificial Intelligence,* **26,** 323–357.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Greene, S. B. (1992). Multiple explanations for multiply quantified sentences: Are multiple models necessary? *Psychological Review,* **99,** 184–187.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Ed.), *Studies in syntax* (pp. 41–58). New York: Academic Press.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology,* **73,** 407–420.

Griggs, R. A., & Newstead, S. E. (1982). The role of problem structure in a deductive reasoning task. *Journal of Experimental Psychology,* **8,** 297–307.

Guyote, M. J., & Sternberg, R. J. (1981). A transitive chain theory of syllogistic reasoning. *Cognitive Psychology,* **13,** 461–525.

Henle, M., & Michael, M. (1956). The influence of attitudes on syllogistic reasoning. *Journal of Social Psychology,* **44,** 115–127.

Henle, M. (1962). On the relation between logic and thinking. *Psychological Review,* **69,** 366–378.

Hintikka, J. (1973). *Logic, language games and information.* Oxford, UK: Oxford Univ. Press.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning and discovery.* Cambridge, MA: MIT Press.

Holyoak, K. J., & Spellman, B. A. (1993). Thinking. *Annual Review of Psychology,* **44,** 265–315.

Horn, L. R. (1989). *A natural history of negation.* Chicago: Univ. of Chicago Press.

Horwich, P. (1982). *Probability and evidence.* Cambridge: Cambridge Univ. Press.

Howson, C., & Urbach, P. (1989). *Scientific reasoning: The Bayesian approach.* La Salle, IL: Open Court.

Jackson, F. (1982). Two modes of syllogistic reasoning. *Communication Monographs,* **49,** 205–213.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge: Cambridge Univ. Press.

Johnson-Laird, P. N. (1994). Mental models and probabilistic thinking. *Cognition,* **50,** 189–209.

Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition,* **16,** 1–62.

Johnson-Laird, P. N., & Byrne, R. M. J. (1989). Only reasoning. *Memory and Language,* **28,** 313–330.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* Hillsdale, NJ: Erlbaum.

Johnson-Laird, P. N., & Byrne, R. M. J. (1990). Meta-logical reasoning: Knights, knaves and Rips. *Cognition,* **36,** 69–84.

Johnson-Laird, P. N., & Byrne, R. M. J. (1992). Modal reasoning, models, and Manktelow and Over. *Cognition,* **43,** 173–182.

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., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J.-P. (in press). Naive probability: A mental model theory of extensional reasoning. *Psychological Review.*

Johnson-Laird, P. N., & Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology,* **10,** 64–99.

Kant, I. (1961). *Critique of pure reason.* (F. M. Muller, Trans.). New York: Doubleday, Dolphin Books.

Kirby, K. N. (1994). Probabilities and utilities of fictional outcomes in Wason's four-card selection task. *Cognition,* **51,** 1–28.

Klauer, K. C. (in press). On the normative justification for information gain in Wason's selection task. *Psychological Review*.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review,* **94,** 211–228.

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

Kosko, B. (1992). *Neural networks and fuzzy systems: A dynamical systems approach to machine intelligence.* Englewood Cliffs, NJ: Prentice Hall.

Levinson, S. (1983). *Pragmatics.* Cambridge: Cambridge Univ. Press.

Lewis, D. (1973). *Counterfactuals.* Oxford, UK: Oxford Univ. Press.

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review,* **85,** 297–315.

Mackie, J. L. (1963). The paradox of confirmation. *British Journal for the Philosophy of Science,* **38,** 265–277.

Manktelow, K. I., Sutherland, E. J., & Over, D. E. (1995). Probabilistic factors in deontic reasoning. *Thinking and Reasoning,* **1,** 201–220.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

McCarthy, J. M., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine intelligence, volume 4.* New York: Elsevier.

McKenzie, C. (1994). The accuracy of intuitive judgment strategies: Covariation assessment and Bayesian inference. *Cognitive Psychology,* **26,** 209–239.

Morgan, J. I. B., & Morton, J. T. (1944). The distortions of syllogistic reasoning produced by personal connections. *Journal of Social Psychology,* **20,** 39–59.

Moxey, L., & Sanford, A. (1991). *Communicating quantities.* Hove, UK: Erlbaum.

Newell, A. (1981). Reasoning, problem solving and decision processes: The problem space as a fundamental category. In R. Nickerson (Ed.), *Attention and performance* (Vol. 8). Hillsdale, NJ: Erlbaum.

Newstead, S. E. (1989). Interpretational errors in syllogistic reasoning. *Journal of Memory and Language,* **28,** 78–91.

Newstead, S. E., & Griggs, R. A. (1983). Drawing inferences from quantified statements: A study of the square of opposition. *Journal of Verbal Learning and Verbal Behavior,* **22,** 535–546.

Newstead, S. E., Pollard, P., & Riezebos, D. (1987). The effect of set size on the interpretation of quantifiers used in rating scales. *Applied Ergonomics,* **18,** 178–182.

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society (Series A),* **231,** 289–337.

Nickerson, R. S. (1996). Hempel's paradox and Wason's selection task: Logical and psychological problems of confirmation. *Thinking and Reasoning,* **2,** 1–31.

Oaksford, M. (1993). Mental models and the tractability of everyday reasoning. *Behavioral and Brain Sciences,* **16,** 360–361.

Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind and Language,* **6,** 1–38.

Oaksford, M., & Chater, N. (1992). Bounded rationality in taking risks and drawing inferences. *Theory and Psychology,* **2,** 225–230.

Oaksford, M., & Chater, N. (1993). Reasoning theories and bounded rationality. In K. I. Manktelow & D. E. Over (Eds.), *Rationality* (pp. 31–60). London, UK: Routledge.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review,* **101,** 608–631.

Oaksford, M., & Chater, N. (1995a). Information gain explains relevance which explains the selection task. *Cognition,* **57,** 97–108.

Oaksford, M., & Chater, N. (1995b). Theories of reasoning and the computational explanation of everyday inference. *Thinking and Reasoning,* **1,** 121–152.

Oaksford, M., & Chater, N. (1995c). Two and three stage models of deontic reasoning. *Thinking and Reasoning,* **1,** 350–357.

Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review,* **103,** 381–391.

Oaksford, M., & Chater, N. (1998a). *Rationality in an uncertain world: Essays on the cognitive science of human reasoning.* Hove, Sussex, UK: Psychology Press.

Oaksford, M., & Chater, N. (1998b). A revised rational analysis of the selection task: Exceptions and sequential sampling. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition.* Oxford, UK: Oxford Univ. Press.

Oaksford, M., Chater, N., Grainger, B., & Larkin, J. (1997). Optimal data selection in the reduced array selection task (RAST). *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **23,** 441–458.

Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory & Cognition,* **18,** 835–854.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems.* San Mateo, CA: Morgan Kaufman.

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.-M. Caverni, J.-M. Fabre, & M. Gonzalez (Eds.), *Cognitive biases: Advances in psychology* (Vol. 68). Amsterdam, Netherlands: North-Holland.

Polk, T. A., & Newell, A. (1988). Modelling human syllogistic reasoning in Soar. In *Proceedings of the Cognitive Science Society–1988* (pp. 181–187). Hillsdale, NJ: Erlbaum.

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review,* **102,** 533–566.

Pollard, P. (1985). Nonindependence of selections on the Wason selection task. *Bulletin of the Psychonomic Society,* **23,** 317–320.

Pylyshyn, Z. W. (Ed.). (1987). *The robot's dilemma: The frame problem in artificial intelligence.* Norwood, NJ: Ablex.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence,* **13,** 81–132.

Reiter, R. (1985). On reasoning by default. In R. Brachman & H. Levesque (Eds.), *Readings in knowledge representation.* Los Altos, CA: Morgan Kaufman.

Revlis, R. (1975a). Two models of syllogistic inference: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behavior,* **14,** 180–195.

Revlis, R. (1975b). Syllogistic reasoning: Logical decisions from a complex data base. In R. J. Falmagne (Ed.), *Reasoning, representation and process.* New York: John Wiley.

Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review,* **90,** 38–71.

Rips, L. J. (1990). Reasoning. *Annual Review of Psychology,* **41,** 321–353.

Rips, L. J. (1994). *The psychology of proof.* Cambridge, MA: MIT Press.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Rumelhart, D. E. (1989). Towards a microstructural account of human reasoning. In S. Vosnaidou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 298–312). Cambridge: Cambridge Univ. Press.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 2: Psychological and biological models* (pp. 7–57). Cambridge, MA: MIT Press.

Russell, B. (1927). *An outline of philosophy.* London, UK: Allen and Unwin.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication.* Urbana: Univ. of Illinois Press.

Siegel, S., & Castellan, N. J. (1988). *Non-parametric statistics for the behavioral sciences.* New York: McGraw Hill.

Shortliffe, E. H., & Buchanan, B. G. (1990). A model of inexact reasoning in medicine. In G. Shafer & J. Pearl (Eds.), *Uncertain reasoning* (pp. 259–273). San Mateo, CA: Morgan Kaufman.

Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory.* Oxford, UK: Oxford Univ. Press.

Stenning, K., & Oaksford, M. (1993). Rational reasoning and human implementations of logics. In K. I. Manktelow & D. E. Over (Eds.), *Rationality* (pp. 136–177). London, UK: Routledge.

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.

Thagard, P. (1988). *Computational philosophy of science.* Cambridge, MA: MIT Press.

Wason, P. C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology,* Harmondsworth, Middlesex, UK: Penguin.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology,* **20,** 273–281.

Wason, P. C., & Brooks, P. G. (1979). THOG: The anatomy of a problem. *Psychological Research,* **41,** 79–90.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology,* **23,** 63–71.

Wetherick, N. E. (1989). Psychology and syllogistic reasoning. *Philosophical Psychology,* **2,** 111–124.

Wetherick, N. E. (1993). Human rationality. In K. I. Manktelow & D. E. Over (Eds.), *Rationality* (pp. 83–109). London, UK: Routledge.

Wetherick, N. E., & Gilhooly, K. (1990). Syllogistic reasoning: Effects of premise order. In K. Gilhooly, M. T. G. Keane, R. Logie, & G. Erdos (Eds.), *Lines of thought, Vol. 1.* London, UK: John Wiley.

Whitehead, A. N., & Russell, B. (1910). *Principia mathematica.* Cambridge: Cambridge Univ. Press.

Wiener, N. (1948). *Cybernetics.* New York: Wiley.

Wilkins, M. C. (1928). The effect of changed material on the ability to do formal syllogistic reasoning. *Archives of Psychology,* **No. 102.**

Wolf, F. M. (1986). *Meta-analysis. Quantitative methods for research synthesis.* Newbury Park, CA: Sage.

Woodworth, R. S., & Sells, S. B. (1935). An atmosphere effect in syllogistic reasoning. *Journal of Experimental Psychology,* **18,** 451–460.

Zadeh, L. (1965). Fuzzy sets. *Information and Control,* **8,** 338–353.

Accepted August 27, 1998