JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 20, 306–321 (1981)

Memory for Logical Quantifiers

JOHN R. ANDERSON

Carnegie–Mellon University

Two experiments are reported on memory for sentences involving the categorical quantifiers *all, some, some not,* and *no.* Three models are considered for quantifier memory: abstract propositional, verbal string, and abstract analog. The first experiment looked at subjects' ability to recognize studied sentences when these study sentences are mixed with distractors that vary in quantifier, underlying relationship, and word order. This experiment produced evidence that quantifier memory depends on both propositional and verbal-string information but not analog information. The second experiment looked at speed to judge quantified statements as true or false on the basis of previously memorized statements. The analog hypothesis predicted that subjects should be fastest to accept *some* statements having studied *all* and to accept *some not* statements having studied *no*. In contrast, it was found that subjects were fastest when study and test quantifiers matched. The experiments also indicated that these quantifiers are ordered in complexity of representation: all < some < no < some not, and that positive and negative sentences are represented quite distinctly. These results are consistent with the ACT propositional representations proposed.

This paper reports two experiments designed to evaluate various proposals for how subjects represent and process logical quantifiers in sentences. While the conclusions are intended to apply to a broad range of sentence quantification, this research uses the four quantifiers that have been important in classical logic (e.g., Cohen & Nagle, 1934) and in the work on categorical syllogisms (e.g., Woodworth & Sells, 1935; Chapman & Chapman, 1959; Erickson, 1974). These four types and examples are:

> Universal Affirmative—All doctors hate lawyers.
>
> Particular Affirmative—Some doctors hate lawyers.
>
> Particular Negative—Some doctors do not hate lawyers.
>
> Universal Negative—No doctors hate lawyers.

Sentences with only one quantified noun were used because they are easy to comprehend relative to more complex quantificational statements like *All doctors do not hate some lawyers.*

Logical quantification has been a central concern in many recent proposals for memory representations (e.g., Anderson, 1976; Anderson & Bower, 1973; Frederiksen, 1975; Kintsch, 1974; Norman & Rumelhart, 1975; Quillian, 1969). These models involve an abstract propositional code for representing information in a form that is not modality-specific. A major purpose of the research reported in this paper is to see whether quantificational information is propositionally encoded as envisioned in these models.

The experiments are designed to distinguish among three major types of memory representations for quantifiers—verbal string, abstract propositional, and abstract analog. At the outset two qualifications should be made to avoid potential misunderstandings of what is being done here. First, the tests will be of three specific theories of representation. It is not possible to test all possible representations of some vaguely specified type. Therefore, a nega-

---

This research was supported by grants BNS76-00959 and BNS78-17463 from the National Science Foundation. I would like to thank Cyndie Grimaldi who ran both of the experiments, Dick Proudfoot who programmed the second experiment, and Lynne Reder and Miriam Schustack who provided comments on the manuscript. Requests for reprints should be addressed to Dr. J. R. Anderson, Department of Psychology, Carnegie–Mellon University, Pittsburgh, Pa. 15213.

0022-5371/81/030306-16$02.00/0

tive conclusion about a particular theory of representation should be interpreted as specific to the theory as formulated in the paper and should not generalize to all possible theories of that kind. Still, the versions of the theories proposed are plausible and are at least similar to ones to which some theorists appear to ascribe. Therefore, conclusions about the representational theories will not be trivial. Moreover, if there is positive evidence for a particular theory as opposed to other possible theories, that is significant support for the theory in that it was supported over plausible alternatives.

As a second qualification, the theories of quantifier representation are theories about how information about quantifiers is *stored and retrieved* as well as about their representation. Thus, the theories being tested are about representation plus process. In line with arguments made elsewhere (Anderson, 1978) it is not possible to test theories of representation without commitment to process assumptions.

Verbal Strings

The most straightforward representation for a quantified sentence is the sentence itself. Exactly what is meant by the "sentence itself" is a little ambiguous. I choose to regard it as a data representation which is a string or sequence of words, where the words themselves are units without any interesting structure for the tasks at hand. This is the type of representation which was proposed by Anderson and Paulson (1977) who found clear experimental evidence for this type of representation. For the purposes of the current experiments, this verbal-string representation is similar to the verbal representations contained in Paivio (1971), Graesser and Mandler (1975), Hayes-Roth and Hayes-Roth (1977), and Kolers (1979). However, I do not mean to assert that these theories are committed to the predictions derived here.

Particularly important to the experiments at hand are the retrieval processes by which subjects determine whether a presented sentence is identical to or implied by a sentence stored in memory. I will assume that this is accomplished by a template matching of the phrases of the presented sentence to phrases of the stored sentence. For present purposes, a template-matching process is one that simply computes a dichotomous result—match or no match. Again, Anderson and Paulson found evidence for such a template-matching process and this process seems compatible with the other verbal-string formulations.

From this description it should be clear that the key feature of the verbal-string model is that it expects memory for quantifiers to be intimately tied to memory for word order. This will contrast with the other two theories which assume that the representation of quantificational information is abstracted away from the particular sentence by which it was communicated.

Propositional Representation

For sake of preciseness, the version of the propositional theory developed for this task will be ACT (Anderson, 1976). However, it should not be thought that conclusions about this alternative are specific to ACT. Other theorists should be consulted for the definitive interpretations of their theories, but it does seem to me that a number of other theories make the same predictions for this paradigm.

Figure 1 presents the ACT network representations for the four types of sentences that will be studied. Clearly, these representations are somewhat removed from the original sentences. Moreover, these representations are abstract in that multiple sentences can result in the same representation. For instance, the representation in Fig. 1a would result whether the subject studied

(1) all doctors hate lawyers; or

(2) lawyers are hated by all doctors.

The representations in Fig. 1 should be briefly explained. The subject–predicate construction in Fig. 1a encodes the fact that the subject, doctor, is a subset (not neces-

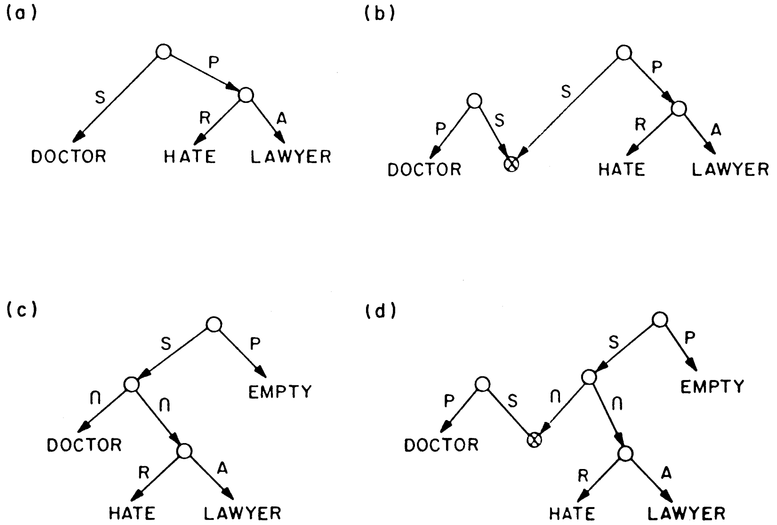

FIG. 1. The ACT representation of the sentences: a. All doctors hate lawyers; b. Some doctors hate lawyers; c. No doctors hate lawyers; d. Some doctors do not hate lawyers.

sarily a proper subset) of the predicate, the people who hate lawyers. In Part (b) of Fig. 1 we have a set X which is a subset of both doctors and those who hate lawyers. Part (c) encodes that the intersection between doctors and those who hate lawyers is empty. The ∩ links denote an intersection between the two sets. Part (d) encodes that the intersection between a subset of doctors and those who hate lawyers is empty.[1]

The ACT model assumes a template-matching process like that of the verbal-string model. However, in this case the templates being matched are network propositions, abstractions from the sentence input. If the representations in Fig. 1 are matched proposition by proposition, we might expect some degree of confusion between (a) and (b) and between (c) and (d). Embedded in (b) is a proposition which almost matches (a), and similarly embedded in (d) is a proposition which almost matches (c). This is a pattern of confusion not predicted under the word string model. If there were any confusions under the word string interpretation it would be between the *some* and *some not* sentences.

An important difference between the abstract-propositional model and the verbal-string model concerns whether subjects should be able to remember the meaning of the sentence even if they cannot remember exact wording. A long series of experiments (e.g., Anderson & Paulson, 1977; Bransford, Barclay, & Franks, 1972; Graesser & Mandler, 1975; Kolers, 1979; Sachs, 1967; Thorndyke, 1976; Wanner, 1968) has now been devoted to exploring this issue. It seems a fair summary that subjects can remember meaning independent of word order—although there is evidence that subjects also remember word order. However, this research has never addressed the issue of whether information as abstract as quantifiers is propositionally stored. There is an alternate interpretation of the meaning-without-exact-wording results which attributes these results to mem-

---

[1] It may seem natural to read Fig. 1a as "The doctor hates the lawyer," but according to the ACT semantics there is a great deal of difference. To encode "the doctor," a structure like Fig. 1b would be required, where the subject of the "hate lawyer" predicate is a subset (node X) of *doctors* rather than the entire doctor set. In addition to the structure in Fig. 1b, the node X would have to be tagged as a singleton set. Another way of encoding "Some doctors do not hate lawyers" would be to tag the propositional structure in Fig. 1a as false. This could then be read as "It is false that all doctors hate lawyers" which is semantically equivalent to the *some not* sentence. It would also be possible to encode the particular negative by negating the main proposition in Fig. 1b. However, we choose Fig. 1d because it used the same convention as Fig. 1c. There is no other way to encode the universal negative in ACT than Fig. 1c.

ory for visual images (Begg & Paivio, 1969; Paivio, 1971). One cannot really develop concrete visual images of the meaning of abstract quantifiers because a true interpretation of their meaning would require representing a potentially infinite universe of objects being quantified over and concrete images cannot represent infinity.

In trying to discriminate the abstract-meaning representations from the verbal-string representation we will be particularly interested in quantifier memory in the situation where the subject can remember the meaning but not the word order. If quantifier memory is greater than chance in this situation, this would be evidence that memory for quantifiers, like other aspects of sentence meaning, can rest on something other than the verbal string.[2]

Abstract Analog

I first came upon the abstract-analog representation for quantifiers in trying to imagine how the Begg and Paivio visual image analysis of sentence meaning would extend to memory for quantifiers. The idea is somewhat removed from anything explicitly developed in the Paivio dual-code theory; nonetheless, there have appeared in the literature a number of proposals by other theorists which are slight variations or instantiations of my proposal.

The basic idea derives from the fundamental difference between analog and symbolic representations. This is that an analog representation mirrors (or analogizes) the meaning of what it is representing in its internal structure, whereas a symbolic representation has an arbitrary relationship to the meaning of what it is representing. According to this distinction, both the verbal string and the propositional representations are symbolic.

A difficulty for analog representations is that there is no way to create a true analog of the abstract meaning of quantifiers. This is because quantifiers refer to a potentially infinite universe while analogs, by virtue of the fact that they are physical models, must be finite. Moreover, a concrete model must make commitments to things that quantifiers are ambiguous about (for instance, a concrete model must choose between representing *some* as *all* or *some but not all*, whereas the abstract proposition does not). Still it is possible to create a partial analog. The initial proposal I had in mind was that subjects commit to memory an image which depicts the relation between subjects and objects. This image can be concrete (e.g., a room full of doctors glaring at lawyers to instantiate *All doctors hate lawyers*) or more abstract (e.g., the Venn diagram for doctors including the Venn diagram for lawyers or for lawyer-haters). In this image subjects try to represent the quantificational information in the sentence. If the quantifier is *all*, every individual in the image possesses the relationship. If the quantifier is *some*, a proper subset does. If the quantifier is *some not*, a proper subset does not, and if the quantifier is *no*, none of the individuals will. We assume that the subject inspects these images at recall to see which of the various quantifiers applies.

It should be noted that many instantiations of this analog model have been proposed for different tasks. For instance, Venn diagrams (Erickson, 1974) are such a representation. Johnson-Laird and Steedman (1978) offer a model as concrete as doctors staring at lawyers, but they also offer a more abstract equivalent representation which has this analog property. Guyote and Sternberg (Note 1) also offer a similar abstract-analog representation. Meyer's (1970) representation for semantic

---

[2] This is not to assert that an advocate of pure verbal string representation could not maintain that position. It is possible to propose "sophisticated guessing strategies" that would reconstruct the meaning and quantifier from string fragments, but not be able to reconstruct exact wording. This takes us back to the earlier remarks about the impossibility of rejecting a whole class of theories based on some representation. We know in principle (Anderson, 1978) that there will be some choice of process which can predict the phenomena. In absence of the possibility of proving false all theories based on a representation, we have to content ourselves with rejecting some and leaving the remainder to judgments of inherent plausibility.

memory judgments also had this same basic character.

Each of these models, and my interpretation here, assumes that subjects verify sentences by inspecting images. To verify *all,* they must inspect whether all the members of the image have the relationship; to verify *some,* they must find some member of the image that does; to verify *some not,* they must find some member that does not; and to verify *no,* they must inspect all members to make sure that none do. This inspection process is quite different from the template-matching process assumed in the verbal-string and propositional models. Rather than matching a pattern against memory and getting an all-or-none result, one inspects pieces of the memory and keeps track of the relationship exemplified by each piece. Note in the template-matching scheme two structures are created, one for the memory and one for the test probe, and the two are matched against each other. In the analog scheme the test quantifier specifies a procedure to be performed on the memory representation. This difference in inspection processes is the source of some of the predictions that differentiate this theory from the other two.[3]

Recently, Holyoak and Glass (1978) have come up with an alternative analog interpretation of the quantifiers *all, most, some, few,* and *none*. They propose that these quantifiers are ordered on a unidimensional continuum of magnitude. They do not treat *some not* but presumably it would be ordered between *some* and *none*. They propose that subjects recognize quantified sentences by making various kinds of magnitude judgments. For instance, to verify *all* subjects try to see whether the stored magnitude is similar to a prototype for *all*. Although this analog model is clearly different from the previous, it turns out to yield similar predictions for the experiments at hand.

In particular, both analog models predict similar though not identical patterns of confusion among quantifiers. Both predict a great deal of confusion between *some* and *some not* in distinction to the propositional model. The first analog image model predicts that subjects having studied *all* would falsely accept *some* because some is true of an *all* image. However, it would not predict confusions in the other direction. The magnitude comparison model would predict confusions in both directions. Similarly, the image model would predict acceptance of *some not* sentences when *none* was studied (but not vice versa) while the magnitude judgment model would predict symmetrical confusions.

EXPERIMENT 1

The first experiment employed a recognition–memory paradigm. Subjects studied facts like

All doctors hate lawyers,

and after studying a set of such sentences, they would see in random order eight sentences like the following:

RQO All doctors hate lawyers.
$RQ\bar{O}$ Lawyers are hated by all doctors.
$R\bar{Q}O$ Some doctors hate lawyers.
$R\bar{Q}\bar{O}$ Lawyers are hated by some doctors.
$\bar{R}QO$ Doctors are hated by all lawyers.
$\bar{R}Q\bar{O}$ All lawyers hate doctors.
$\bar{R}\bar{Q}O$ Doctors are hated by some lawyers.
$\bar{R}\bar{Q}\bar{O}$ Some lawyers hate doctors.

Their task was to indicate which of these sentences they had studied. The three-letter prefix before each probe (which was not presented to subjects) indicates what information is preserved in the probe sentence: R indicates subject–verb–object semantic relation, Q indicates quantifier, and O indicates order of subject and object.

---

[3] One could propose a template-matching process in which one analog was matched against another. This would produce very similar predictions to the verbal string and propositional models. However, this process seems in violation of the spirit of an analog model. The point of an analog model is to incorporate in the analog structure the knowledge being represented. This seems a meaningful point only if the processes that use the analog interpret the information encoded in the structure. Otherwise, the analog is being processed symbolically—as a representation whose structure might as well be arbitrary with respect to what is being represented.

A bar over the letter indicates that this information is not preserved in the probe. By looking at false alarms to various of these choices we can diagnose what the subject remembers. By looking at the pattern of false alarms to alternative quantifiers, we can address the question of affinities among the quantifiers.

Method

Subjects. Forty undergraduates were recruited from the student population at Yale. They participated in the 1-hour experiment either for $2.50 or to receive credit for an undergraduate psychology course.

Materials. A study list of 48 sentences was randomly created for each subject. Half of the sentences (24) involved transitive verbs and the other half involved comparative adjectives. In the comparative case, subjects studied sentences like

> Some doctors are not smarter than lawyers.

For this particular target, the eight sentences at recognition test might be:

RQO Some doctors are not smarter than lawyers.
$RQ\bar{O}$ Lawyers are not dumber than some doctors.
$R\bar{Q}O$ All doctors are smarter than lawyers.
$R\bar{Q}\bar{O}$ Lawyers are dumber than all doctors.
$\bar{R}QO$ Doctors are not dumber than some lawyers.
$\bar{R}Q\bar{O}$ Some lawyers are not smarter than doctors.
$\bar{R}\bar{Q}O$ Doctors are dumber than all lawyers.
$\bar{R}\bar{Q}\bar{O}$ All lawyers are smarter than doctors.

Half of the transitive-verb sentences (12) were active and half were passive, with each sentence involving a different verb. The comparative sentences were also divided into two halves, such that opposite comparatives appeared in each half; one half had unmarked comparatives (e.g., taller) and the other half marked comparatives (e.g., shorter). Thus, the material consisted of four sets of 12 sentences. For each set of sentences, 3 were presented with each of the four quantifiers (*all, some, some not,* and *no*). Thus there were 16 *subsets* of 3 sentences. Sentence subjects and objects were randomly assigned to these sentences from a master list of 96 nouns, such that each experimental subject had a different set of sentences. For each study sentence there was a test set of eight alternatives. Each sentence of a subset of 3 sentences was assigned to a different foil quantifier (there were three other quantifiers possible beside the one in the studied sentence). The foil quantifier is the other quantifier that is used in the set of 8 test sentences for the study sentence. The assignment of foil quantifier to sentence was random, as was the ordering of the eight alternatives.

Procedure. The entire experiment was run by computer and all material was presented via a CRT screen. The computer system used is described by Proudfoot (1978). The study sentences were presented individually on the screen in random order. For each sentence, the subject was prompted after 5 seconds to give a rating on a 1–7 scale according to how well the sentence described the true state of affairs in the world. Thus, the subject studied each sentence for at least 5 seconds but total study time (which was recorded) was under the subject's control. Subjects were not told that this was a memory task. The truth-rating task was used to encourage the subjects to process the material in a meaningful way. The incidental, meaning-orienting paradigm was used to maximize the probability that subjects would use an abstract-propositional representation if they could. The purpose of this experiment was not to find out the range of ways subjects might encode quantifiers, but rather to see if subjects were capable of propositional encodings of quantifiers.

After the study phase, subjects were informed that this was a memory experiment and were presented with sets of eight alternatives for each studied sentence. The subjects' task was to indicate the exact sentence studied. The eight alternatives were presented on the screen in random order numbered 1–8. The subject typed in the number which indicated his choice. The sentences were tested in the same order as studied, in order to keep a relatively constant lag.

Results

Because of a computer error study time data for 14 of the subjects were lost. The study time data for the remaining 26 subjects are presented in Table 1 broken down according to whether the sentence was transitive or comparative and according to quantifier. Within the transitive sentences, the data are further broken down according to whether the sentence is active or passive. Within the comparative sentences the data are broken down according to whether a marked or unmarked adjective is used. A within-subjects analysis of variance was performed and all main effects were significant: transitive vs comparative ($F(1,25) = 12.85$; $p < .001$); quantifier ($F(3,75) = 16.28$; $p < .001$); active vs passive ($F(1,25) = 4.65$; $p < .05$); unmarked vs marked ($F(1,25) = 24.96$; $p < .001$). An interpretation of the different study times for different quantifiers will be offered at the end of the paper.

There was only one significant interaction—between quantifier and sentence type. This can be explained by observing that the differences among sentence types increases for more difficult quantifiers. Consider the extreme difference between active transitive sentences and marked comparatives. This difference is 0.81 seconds for *all,* 1.36 for *some,* 2.56 for *some not,* and 3.75 for *no.* The variation among these differences is quite significant ($F(3,75) = 7.20$; $p < .001$).

TABLE 1

STUDY TIME (sec) AS A FUNCTION OF SENTENCE TYPE

	Transitive sentences				
	All	Some	Some not	No	Mean
Active	9.71	10.73	11.12	10.76	10.46
Passive	9.99	10.78	12.15	11.81	11.18
Mean	9.85	10.76	11.64	11.04	10.82
	Comparative sentences				
	All	Some	Some not	No	Mean
Unmarked	10.12	10.27	11.81	11.04	10.74
Marked	10.52	12.06	13.68	14.51	12.69
Mean	10.32	11.17	12.60	12.78	11.72
Overall means	10.09	10.97	12.12	11.91	

Table 2 presents the data on proportion of correct and incorrect responses, classified according to whether the sentence used a transitive verb or a comparative, and according to whether relation (R), quantifier (Q), and order (O) were correctly identified in the recognition choice. Also we give the probabilities (collapsed over the other dimensions) of subjects' being correct as to relation, quantifier, and word order for the two types of material (transitive and comparatives). All six of these probabilities are significantly greater than chance (that is, .50) by statistical tests computed with subjects as a random variable (by t tests, at $p < .001$ or better). In particular, memory for order was greater than chance, indicating that subjects have some memory for word string information in this situation.

The transitive-verb data differ somewhat from the comparative-adjective data. Subjects show the best memory for the relation in transitive-verb sentences but actually best memory for the noun order with the comparatives. This rather good memory for order relative to meaning for comparatives has also been reported by Tieman (1971).

Table 2 also reports all the conditional probabilities for recognition of quantifiers. Of particular importance is the probability conditional on recall of relation but not order, $P(Q|R\bar{O})$. These conditional probabilities are both greater than chance (.5); the difference is very significant ($t(39) = 7.46$; $p < .001$) for the transitive verbs and marginally significant ($t(39) = 1.92$; $p < .05$, 1-tailed) for the comparatives. It is also of interest to note that relation is more predictive of quantifier recall than is order. This is seen most clearly in the fact that

TABLE 2

PROPORTION OF CHOICES AMONG RECOGNITION ALTERNATIVES

	Transitive verbs				
	O			$\bar{O}$	
	Q	$\bar{Q}$		Q	$\bar{Q}$
R	.433	.124	R	.164	.070
$\bar{R}$	.051	.043	$\bar{R}$	.063	.051
$P(R) = .791$	$P(Q) = .711$		$P(O) = .651$		
Conditional analyses					
$P(Q \mid R) = .755$	$P(Q \mid \bar{R}) = .548$		$P(Q \mid O) = .743$	$P(Q \mid \bar{O}) = .652$	
$P(Q \mid RO) = .777$	$P(Q \mid R\bar{O}) = .701$		$P(Q \mid \bar{R}O) = .543$	$P(Q \mid \bar{R}\bar{O}) = .553$	
	Comparative adjectives				
	O			$\bar{O}$	
	Q	$\bar{Q}$		Q	$\bar{Q}$
R	.341	.141	R	.089	.066
$\bar{R}$	.110	.106	$\bar{R}$	.078	.071
$P(R) = .637$	$P(Q) = .618$		$P(O) = .698$		
Conditional analyses					
$P(Q \mid R) = .675$	$P(Q \mid \bar{R}) = .515$		$P(Q \mid O) = .646$	$P(Q \mid \bar{O}) = .549$	
$P(Q \mid RO) = .707$	$P(Q \mid R\bar{O}) = .574$		$P(Q \mid \bar{R}O) = .509$	$P(Q \mid \bar{R}\bar{O}) = .523$	

Note. R = correct meaning relation, $\bar{R}$ = reversed meaning relation, Q = correct quantifier, $\bar{Q}$ = foil quantifier, O = correct noun order, $\bar{O}$ = reversed noun order.

$P(Q|R\bar{O})$ is .112 larger than $P(Q|\bar{R}O)$. This difference is quite significant ($t(39) = 3.17$; $p < .001$). The propositional and analog models developed in the introduction, but not the verbal-string model, predicted better than chance memory for quantifiers in the $R\bar{O}$ state.

The difference in conditional quantifier recall between the RO state and the $R\bar{O}$ state is a measure of the contribution of word order (and hence verbal strings) to quantifier memory. This difference is .104 and is quite significant ($t(39) = 2.52$; $p < .01$). Thus, it seems clear that verbal-string memory is also important to quantifier memory.

Table 3 presents two confusion matrices suitable for discriminating between the abstract-propositional model and the symbolic-imagery model. One matrix (a) presents the confusions for the total data and the other (b) the probabilities for the data conditional on correct recall of relation. These matrices list by rows the actual quantifier studied and by columns possible foil quantifiers. The proportions entered are probabilities of incorrectly accepting each foil quantifier. Hits for a particular study-test foil combination are just 1 minus the false alarms for that cell. There are X's down the main diagonal because, of course, a study quantifier could not be used as its own test foil. The critical question is whether there is any special affinity for certain foils given certain originals. It is true that the cells in Table 3 are not all equal but this may only reflect poorer memory for a quantifier or a greater response bias toward a quantifier.

At the bottom of Table 3 we have summarized more focused efforts to look for quantifier affinities in a way that gets rid of the missing cells and makes it easy to subtract out effects of differential study or differential response bias. Part (c) presents a collapsed version of Table (a) which classifies errors according to whether the erroneously accepted test sentence involved a positive term (some, all) or a negative term (some-not, no) and whether the originally

TABLE 3

False Alarm Rates (Proportions) as a Function of Quantifier Studied and Quantifier Tested

(a) Total Data

Quantifier studied	Foil quantifier at test: All	Some	Some not	No
All	X	.399	.357	.408
Some	.267	X	.294	.384
Some not	.234	.357	X	.414
No	.240	.294	.309	X

(b) Conditional on recall of relation

Quantifier studied	Foil quantifier at test: All	Some	Some not	No
All	X	.339	.273	.294
Some	.264	X	.240	.312
Some not	.207	.303	X	.435
No	.186	.288	.277	X

(c) Analysis of bias for logical sign—total data

	Positive	Negative
Positive	.333	.361
Negative	.281	.362
Interaction	+.053	

(d) Analysis of bias for logical sign—conditional data

	Positive	Negative
Positive	.302	.280
Negative	.246	.356
Interaction	+.132	

(e) Analysis for confusion among particular sentences—total data

	Particular	Distant universal
Particular	.326	.309
Distant universal	.326	.324
Interaction	+.015	

(f) Analysis for confusion among particular sentences—conditional data

	Particular	Distant universal
Particular	.272	.260
Distant universal	.281	.240
Interaction	−.029	

studied sentence involved a positive or negative term. Part (d) presents the same condensation for the data in Part (b). If subjects tended to make false alarms in terms of the sign of the logical term we would expect an interaction in the data with the on-diagonal cells larger than the off diagonal. This is just what is found with the larger interaction occurring in the data conditional in recall of meaning. The interaction that appears in matrix (d) is quite significant ($t(39) = 2.43, p < .01$).

Both the propositional model and the analog models predict the interaction that appears in matrix (d). The propositional model predicts the interaction becaues of the similarity in the structures that encode positives (parts a and b of Fig. 1) and the similarity in the structures that encode negatives (parts c and d of Fig. 1). The analog model predicts the interaction because these quantifiers have similar analog encodings. However, analog models also predict the most confusion between *some* and *some not* because they tend to have very similar encodings whereas the propositional model does not. This prediction is tested in matrices (e) and (f). In matrix (e) we have the data classified according to whether a particular quantifier was used at study or test and whether a distant universal was used. A distant study universal is defined as a universal that is not adjacent to the test quantifier on the scale *all, some, some not,* and *no*. Similarly, a distant test universal is a universal quantifier at test not adjacent to the study quantifier on the same scale. So, missing from part (e) are the cases of particulars combined with adjacent universals (i.e., the following four cells from part (a); study all, test some; study some, test all; study some not, test no; study no, test some not). The analog theory predicts high confusions in these particular–universal cases. This exclusion also serves to keep equal the relative frequencies of each quantifier down columns or across rows in Part (e). Part (f) was derived from Part (b) in the same way that Part (e) was derived from Part (a). The abstract-analog model predicts that the particular–particular cell should be much higher than predicted from row or column averages. This amounts to a prediction of interaction in the matrices. As can be seen the particular–particular cells are not high and there is no evidence for the predicted interaction.

Thus we see that subjects can remember whether the sentence was positive or negative independent of quantifier, but they cannot remember if the sentence was particular or universal independent of quantifier. Recall that the two positive representations in Fig. 1 are quite similar as are the two negative representations. Thus, the propositional model can predict some confusion among same-sign sentences. In contrast, there is little basis in the representations of Fig. 1 for confusion between *some* and *some not* sentences.

It should be pointed out, in this context, that the test for confusion among particular quantifiers in Parts (e) and (f) of Table 3 is potentially quite powerful. Holyoak and Glass (1978) found a high degree of confusion among *many, some,* and *a few,* which were the particular quantifiers they used (rather than *some not*). Had the same test as in Tables 3e and f been performed on their data, there would have been a very strong interaction.

The data reported here cannot be fit by the analog model that Holyoak and Glass successfully fit to their data. So the two sets of data are inconsistent. There are any number of procedural differences between the two experiments which might account for the results. However, the most likely would seem to be the exact quantifiers used. Holyoak and Glass used *all, many, some, a few,* and *none*. The inclusion of *many* and *few* suggests analog representations. The *all, some, some not,* and *no* are much more difficult to represent because the ordering of *some* and *some not* is logically indeterminant. Therefore, it seems that the material in this experiment forced subjects to represent the logical meaning of the quantifiers, while the material of Holyoak and Glass encouraged subjects to treat the terms like intensifiers.

Discussion

It would be useful to review the state of knowledge with respect to models set forth in the introduction. Quantifier recall is highly conditional on recall of relation even when the word order cannot be recalled. This supports the propositional and analog models in which an important component of quantifier memory is a representation of the sentence's meaning. It did not seem that the confusions among quantifiers were in the direction expected by the abstract-analog hypothesis, but rather in the direction predicted by the propositional model. Finally, quantifier recall was more highly conditional on recall of both relation and order than just relation, indicating a significant contribution of word order memory to quantifier memory. These results point to abstract propositions and verbal strings but not analog representations as means of encoding quantifier information.

EXPERIMENT 2

The rejection of analog representations in Experiment 1 rests on failing to reject the null hypothesis in matrices (e) and (f) of Table 3. It would be preferable for rejection of this model to depend on rejecting a statistical hypothesis; this is the function of the second experiment. Subjects committed to memory sentences involving the same quantifiers used in the previous experiment. In contrast to the previous experiment which required exact wording judgments, this experiment required truth judgments. Independent of what quantifier was studied with a sentence frame, subjects could be asked to verify that sentence frame with any quantifier. That is, we had looked at all 16 possible combinations of the four quantifiers at study and at test. Our interest was in what would happen when the subject studied *all* and was tested with *some* or studied *no* and was tested with *some not*. For instance, the subject might study:

All lawyers hate doctors.

and be tested with:

Some lawyers hate doctors.

or study:

No lawyers hate doctors.

and be tested with:

Some lawyers do not hate doctors.

Both of these circumstances involve a positive response that we could compare with other positive-response cases, which involve the same quantifier at test as at study.

The propositional and the verbal-string models predict that subjects should be faster when the quantifier at test matches the study quantifier than when it is only logically implied. In these cases, a simple match would lead to a correct answer. When the test quantifier was a particular and the study quantifier a universal, the subject would have to perform a mental transformation.

In contrast, the analog model predicts fastest times when the quantifier is switched from a universal at study to a particular at test. Recall that a universal is represented as an image of all subjects bearing the asserted relationship to the object. At test a subject inspects the image to determine if the quantifier in the test sentence is true of that image. It would be easier (or at least not harder) to verify from such an image that a subset (*some*) of the individuals bore the relation than to check that *all* did, since fewer individuals have to be checked in the *some* case. This implies that having studied *all, some* should be faster than *all* (or at least as fast). Additionally, it should be easier (or at least not harder) to find a subset bearing the desired relation from an image where all bear the relationship than from an image where only a subset bear the relationship. This implies *some* should be easier to verify having studied *all* than having studied *some*. This set of predictions also follows, for instance, from Meyer's (1970) model for processing analog representations in semantic memory. By similar reasoning, we can conclude that it should be easier to verify *some not* having studied *no* than *no* having studied *no*, or *some not* having studied *some not*.

The analog theory of Holyoak and Glass (1978) was not addressed to this task but it would seem to make the same predictions. To verify a *some* relationship, the subject would have to determine that the stored quantifier was equal to or greater than *some* on the quantifier continuum. This discrimination should be easier if the stored quantifier is *all* than if it is *some*. This prediction derives from the work on comparison of analog quantities (e.g., Moyer, 1973; Holyoak & Walker, 1976). By similar analysis the analog model would predict judging *some not* having studied *no* would be the easiest of the judgments for true negative sentences.

Method

Subjects. Fifteen subjects were recruited from the same population as used in Experiment 1. They participated in the 2 hour experiment either to receive credit in introductory psychology or for $5.00 pay.

Materials. Experiment 1 indicated that subjects displayed better memory for meaning with transitive verbs than with comparatives. Therefore, all sentences studied were of the subject–verb–object variety. Ten sentences involved universal affirmative quantification of the subject, 6 particular affirmative, 6 particular negative, and 10 universal negative. These 32 sentences were randomly constructed for each subject from a set of 32 verbs and 64 nouns referring to people.

Procedure. The total experiment was run by the same computer arrangement as the first. Subjects were given at least 5 seconds to study each sentence. They were told to try to commit the sentences to memory in a meaningful rather than a rote way—to try to understand why the asserted relationship might be true. Five seconds after each sentence appeared subjects were prompted to rate sentence comprehensibility on a 1–7 scale. After they gave this rating, they saw the next sentence. After this study phase, they were transferred to a study-test phase in which they were presented with subject, verb, and object in scrambled order and had to reconstruct the original sentence with quantifier. A double dropout procedure was used. The subject passed through the 32 sentences (in the same order as study) and incorrect responses were noted. Missed sentences were retested until they were correctly reported. The entire testing procedure was then repeated to achieve the second dropout.

A reaction-time test phase followed the study-test phase. Sentences were presented on the screen and the subject had to hit a "k" or "d" on the keyboard to indicate whether the test sentence did or did not logically follow from what had been studied. Responses and times were recorded. The exact meaning of each quantifier was explained to the subjects, as well as which study quantifiers implied which test quantifiers. In particular, subjects were told that *some* quantifiers were true of *all* study sentences and *some not* quantifiers were true of *no* study sentences.

The sentences studied with the various quantifiers were randomly assigned to be tested with any of the four quantifiers. Table 4 illustrates the frequency with which sentences studied with each quantifier were assigned to be tested with the various quantifiers. These frequencies were chosen to assure that each test quantifier was tested equally often as a true and as a false, and so that a study sentence would have equal probabilities of being tested true and false, regardless of its quantifier. This distribution avoids any chance of the subject's biasing his response selectively on the basis of the test or study quantifier.

TABLE 4
DISTRIBUTION OF MATERIALS IN VERIFICATION TESTS

Quantifier studied	Quantifier tested			
	All	Some	Some not	No
All	3	2	4	1
Some	1	3	1	1
Some not	1	1	3	1
No	1	4	2	3

The basic pattern for testing in Table 4 was repeated for 10 blocks, with each study sentence tested once per block. In each block, the assignment of study sentences to test quantifiers was randomized within the frequency constraints of Table 4. Also, order of testing sentences was randomized. This meant that there were between 10 and 40 observations per subject per cell of Table 4. After each 2 blocks, the subject was given the opportunity to take a rest, get up, walk around, and so on.

Results

The concern is with performance in the reaction-time phase. Table 5 indicates the accuracy and reaction times for the various combinations of study and test quantifier. Reaction times are computed only from those trials where the subject's response was correct. The 16 condition-means for each subject went into an analysis of variance with factors of subjects and conditions. The standard error of the mean reaction time in Table 5 was 135 milliseconds computed from the subject × condition interaction. The standard error of mean error rates, similarly computed, is .038. Each standard error has 210 degrees of freedom. The reaction times and error rates are greater when a true particular is tested of a universal than when a true universal is tested of a universal or a true particular is tested of a particular. These are effects in the direction predicted by the abstract-propositional or verbal-string models and opposite to the predictions of the analog models. Three of these four reaction-time differences are significant: (study)all−(test)some > all−all, $t(210) = 2.56, p < .01$; no−some not > no−no, $t(210) = 6.53, p < .001$; and no−some not > some not−some not, $t(210) = 6.03, p < .001$. The differences between the all−some and some−some conditions is not significant ($t(210) = 1.23$, $p > .1$) but is in the direction predicted by the propositional model. Thus, we have a

TABLE 5
MEAN REACTION TIME AND PERCENTAGE ERRORS (IN PARENTHESES) FOR THE CONDITIONS OF EXPERIMENT 2[a]

Quantifier studied	Quantifier tested All	 Some	 Some not	 No
All	1.509^{T}	1.998^{T}	2.386^{F}	1.911^{F}
	(.065)	(.083)	(.163)	(.113)
Some	2.164^{F}	1.764^{T}	2.806^{F}	2.289^{F}
	(.220)	(.049)	(.373)	(.140)
Some not	2.047^{F}	3.063^{F}	1.857^{T}	2.588^{F}
	(.140)	(.460)	(.080)	(.173)
No	1.933^{F}	2.066^{F}	3.008^{T}	1.761^{T}
	(.107)	(.167)	(.277)	(.144)

Reanalysis of false data
Distance

	1			2			3		
	S–A	2.164	(.220)	A–SN	2.386	(.163)	A–N	1.911	(.113)
	S–SN	2.806	(.373)	S–N	2.289	(.140)	N–A	1.933	(.107)
	SN–S	3.063	(.460)	SN–A	2.047	(.140)			
	SN–N	2.588	(.173)	N–S	2.066	(.167)			
Mean		2.655	(.307)		2.197	(.153)		1.922	(.110)

[a] *Note*. The superscript T's indicate conditions that require true responses and F's indicate conditions that require false responses.

clear refutation of the analog model and clear support for the abstract-propositional model.

The data for the false responses in Table 5 display an interesting pattern: The quantifiers can be seen as on a continuum ordered *all, some, some not,* and *no*. The second half of Table 5 presents the false data classified as to distance apart on this continuum. There are four cases of distance 1, four of distance 2, and two of distance 3. The first letter or letters for each condition refers to quantifier studied and the second to quantifier tested. So S–A stands for *some* studied and *all* tested. There is better performance the further the quantifiers are apart. All but 2 of the 32 possible reaction-time comparisons and all 32 error rate comparisons satisfy this trend. Subjects are particularly poor when they must reject a *some,* have studied a *some not,* or vice versa. This general pattern of data is consistent with the analog comparison model. The problem for this model concerns the true judgments.

The false data are partially consistent with the propositional model. All the quick rejections (distances 2 and 3 in Table 5) involve comparison of positive test sentences and negative study sentences or vice versa. Given the earlier discussions of the similarities among same-sign propositonal representations and differences between different-sign representations (see Fig. 1) one would expect these six conditions to be fast. The differences among these six conditions are relatively small and could be largely explained in terms of the relative complexity of quantifiers (*some not* most complex; *all* simplest). The major problem for the propositional explanation is that the pairings of *some* with *some not,* which involve a sign mismatch, result in the poorest performance.

Using a suggestion of a reviewer, it is possible to account for the slow times to the pairings of *some* and *some not* by supposing that subjects made certain pragmatic inferences upon studying these sentences. For instance, upon learning *Some doctors hate* *lawyers* subjects may infer that *Some doctors do not hate lawyers,* reasoning that if this were not so the universal quantifier *all* would have been used. This pragmatic inference follows from Grice's (1967) conversational maxims. Similarly, upon studying *some not,* subjects could make the inference *some*. If subjects were making these inferences, they would have stored in memory inferences matching the test probe. Their long reaction times would reflect having to reject the spurious matches to the pragmatic inferences. The problem with the assumption of pragmatic inferences is that it would seem to predict recognition confusions among *some not* and *some* in Experiment 1, which did not obtain. To explain the lack of confusion in Experiment 1, one must assume that under the relaxed testing conditions of that experiment, subjects had time to sort out the difference between pragmatic inferences and what they initially studied, even if there might be an initial spurious match between the test sentence and the pragmatic inference. Admittedly, all this is stretching the plausibility of a post hoc explanation.

General Conclusions

The first experiment indicated that subjects have fairly good memory for sentence quantifiers—even when sentences are presented just once, relatively briefly, and without supporting context. Subjects showed fairly good memory for the quantifier when they could only recall the subject–verb–object relation and not word order, but they showed even better memory when they could recall the word order as well. These data are consistent with the assumption that subjects were storing and using both propositional and word string representations as described in the introduction.

The two experiments accumulated some evidence against the analog model for quantifier memory. The first experiment failed to find memory confusions between *some* and *some not*—as would be predicted by the analog model. The second experiment found that subjects were slower to judge particulars having studied universals than they were to judge particulars having studied particulars or to judge universals having studied universals. This is predicted by the propositional model but the opposite was predicted by the analog model. The false rejection data in Experiment 2 were ambiguous—providing a pattern of data that could possibly be explained with either model. However, the propositional model offers an explanation consistent with the other results.

It is interesting to inquire as to what these data say about the relative difficulty of the four quantifiers. Using reading times in Table 1 we have *all* 10.09 seconds, *some* 10.97 seconds, *no* 11.91 seconds, and *some not* 12.12 seconds. Using the times from Experiment 2 to verify true assertions when the quantifiers match, the times are *all* 1.51 seconds, *some* 1.76 seconds, *no* 1.76 seconds, and *some not* 1.86 seconds. Although *some* and *no* are almost tied in terms of verification times, there were many more errors for *no* statements, suggesting some speed–accuracy trade-off. So, both experiments using two different measures seem to agree in the order of difficulty: all < some < no < some not. This ordering closely corresponds to the ordering of complexity of the ACT network representations in Fig. 1. In terms of number of links we have all = 4, some = 6, no = 6, and some not = 8. Although *some* and *no* are equal in number of links, *no* should be more complicated than *some* because it involves an extra level of propositional embedding.

A number of other metrics fail to order the four quantifiers. Length of sentence does a poor job because by this criterion *no* sentences should be fastest. There is no reason to expect this ordering on the basis of symbolic imagery or analoglike representations. Standard predicate calculus logics also fail to predict the ordering. Using separate universal and existential quantifiers we would prefix the four propositions for the four types as follows: all = $(\forall x)$; some = $(\exists x)$; no = $(\forall x)\sim$; some not = $(\exists x)\sim$. This fails to predict the fact that existential (or particular) statements are more complex. One could define the existential quantifier in terms of the universal as is done in some variations of standard predicate calculus. Then we would have the following representation: all = $(\forall x)$; some = $\sim(\forall x)\sim$; no = $(\forall x)\sim$; some not = $\sim(\forall x)$. This incorrectly predicts that *some* has the most complex representation. The data indicate that (1) particulars are more complex than universals, (2) negatives are more complex than positives, and (3) the negative effect is greater than the particular effect. While ACT is not the only conceivable representation that could capture this ordering, it is one.

The presupposition in going into this research was that quantifier memory would depend both on memory for the verbatim sentence and on some abstract representation of meaning. This seems the obvious extrapolation of past research on memory for meaning versus exact wording of sentences. The critical question was whether the meaning component of quantifier memory depended upon a propositional or an abstract representation. This critical question was refined into a test between a specific propositional model and two specific analog models. The first experiment attempted to assess the validity of the presupposition as well as the critical question. The second experiment was focused on the critical question. The evidence generally favored the propositional model as involved in the representation of the meaning of logical quantifiers.

REFERENCES

Anderson, J. R. *Language, memory, and thought.* Hillsdale, N.J.: Erlbaum, 1976.

Anderson, J. R., & Bower, G. H. *Human associative memory.* Washington, D.C.: Hemisphere Press, 1973.

Anderson, J. R., & Paulson, R. Representation and retention of verbatim information. *Journal of Verbal Learning and Verbal Behavior,* 1977, **16,** 439–451.

Anderson, J. R., & Paulson, R. Interference in memory for pictorial information. *Cognitive Psychology,* 1978, **10,** 178–202.

Begg, I., & Paivio, A. Concreteness and imagery in sentence memory. *Journal of Verbal Learning and Verbal Behavior,* 1969, **8,** 821–827.

Bransford, J. D., Barclay, J. R., & Franks, J. J. Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology,* 1972, **3,** 193–209.

Chapman, L. J., & Chapman, J. P. Atmosphere effect reexamined. *Journal of Experimental Psychology,* 1959, **58,** 220–226.

Cohen, M. R., & Nagle, E. *An introduction to logic and scientific method.* New York: Harcourt, Brace, 1934.

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

Frederikson, C. H. Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology,* 1975, **7,** 371–458.

Graesser, A., & Mandler, G. Recognition memory for the meaning and surface structure of sentences. *Journal of Experimental Psychology: Human Learning and Memory,* 1975, **104,** 238–248.

Grice, H. P. *Logic and conversation.* The William James Lectures, Harvard University, 1967.

Hayes-Roth, B., & Hayes-Roth, F. The prominence of lexical information in memory representations of meaning. *Journal of Verbal Learning and Verbal Behavior,* 1977, **16,** 119–136.

Holyoak, K. J. Comparative judgments with numerical reference points. *Cognitive Psychology,* 1978, **10,** 203–243.

Holyoak, K. J., & Glass, A. L. Recognition confusions among quantifiers. *Journal of Verbal Learning and Verbal Behavior,* 1978, **17,** 249–264.

Holyoak, K. J., & Walker, J. H. Subjective magnitude information in semantic orderings. *Journal of Verbal Learning and Verbal Behavior,* 1976, **15,** 287–299.

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

Kintsch, W. *The representation of meaning in memory.* Hillsdale, N.J.: Erlbaum, 1974.

Kolers, P. A. A pattern analyzing basis of recogntion. In F. I. M. Craik & L. S. Cermak (Eds.), *Levels of processing in human memory.* Hillsdale, N.J.: Erlbaum, 1979.

Kosslyn, S. M., & Bower, G. H. The role of imagery in sentence memory: A developmental study. *Child Development,* 1974, **45,** 30–38.

MEYER, D. On the representation and retrieval of stored semantic information. *Cognitive Psychology*, 1970, **1**, 242–300.

MOYER, R. S. Comparing objects in memory: evidence suggesting an internal psychophysics. *Perception and Psychophysics*, 1973, **13**, 180–184.

MOYER, R. S., & LANDAUER, T. K. Time required for judgments of numerical inequality. *Nature* (*London*), 1967, **215**, 1519–1520.

NORMAN, D. A., & RUMELHART, D. E. *Explorations in cognition*. San Francisco: Freeman, 1975.

PAIVIO, A. *Imagery and verbal processes*. New York: Holt, Rinehart & Winston, 1971.

PROUDFOOT, R. G. YEPS: Running real-time experiments on a timesharing system. *Behavior Research Methods and Instrumentation*, 1978, **10**, 291–296.

QUILLIAN, M. R. The teachable language comprehender. *Communications of the ACM*, 1969, **12**, 459–476.

SACHS, J. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, 1967, **2**, 437–442.

TIEMAN, D. C. *Recognition memory for comparative sentences*. Unpublished doctoral dissertation, Stanford University, 1971.

THORNDYKE, P. W. The role of inferences in discourse comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1976, **15**, 437–446.

WANNER, H. E. *On remembering, forgetting, and understanding sentences: A study of the deep structure hypothesis*. Unpublished doctoral dissertation, Harvard University, 1968.

WOODWORTH, R. S., & SELLS, S. B. An atmospheric effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, 1935, **18**, 451–460.

REFERENCE NOTES

1. GUYOTE, M. J., & STERNBERG, R. J. *A transitive-chain theory of syllogistic reasoning* (NR150-412 Technical Report No. 5). New Haven, Conn.: Department of Psychology, Yale University, March 1978.
2. ANDERSON, J. R. *Recognition confusions in sentence memory*. Unpublished manuscript, 1972.

(Received December 1979)