Children's Ability to Detect Semantic Contradictions

Arnold L. Glass

Rutgers University

Keith J. Holyoak

University of Michigan

Nancy E. Kossan

Stanford University

GLASS, ARNOLD L.; HOLYOAK, KEITH J.; and KOSSAN, NANCY E. *Children's Ability to Detect Semantic Contradictions.* CHILD DEVELOPMENT, 1977, 48, 279–283. Children in grades 1, 3, and 5 were presented with contradictory sentences such as "Some kings are queens" and asked to decide whether the sentences were true or false. The age at which children first were able to evaluate the false sentences correctly corresponded to the relative speed with which the sentences can be evaluated by adults in a timed verification task. These results suggest that the organization of semantic information in memory is similar for school-age children and adults.

Numerous studies of semantic memory have explored the processes by which adults use information about word meanings to evaluate the veracity of simple sentences such as "A rose is a flower" (Collins & Quillian 1969; Glass, Holyoak, & O'Dell 1974; Loftus 1973; Rips, Shoben, & Smith 1973; Wilkins 1971). In these studies, the subject's latency to indicate whether a sentence is true or false has typically been the main dependent measure, since accuracy is generally very high. An important developmental question is how the semantic organization of the adult develops in the child. Is semantic information acquired by a cumulative process, or is it substantially reorganized at various stages of development? Do children and adults show a similar pattern of difficulty for different types of sentences? To show a link between child and adult performance in sentence-verification tasks, one could provide evidence that differences in adult latencies to verify sentences are paralleled by the order in which children learn to assess correctly the validity of the sentences.

Rosch (1973) has provided such evidence for one class of true sentences. She presented both adults and children (9–11-year-old boys) with true and false sentences asserting that an instance was a member of a category (e.g., "A ball is a toy"). She found that for both age groups those true sentences containing frequently produced instances of the predicate category (from the norms of Battig & Montague [1969]) were verified more quickly than sentences containing less frequent instances (e.g., "A ball is a toy" vs. "A swing is a toy"). The children also had a much lower error rate with the frequent than with the less frequent items. Since these errors were made under time pressure, this result does not necessarily mean that the children sometimes did not know that the less frequent instances were members of the predicate category. Nevertheless, Rosch's data are consistent with the hypothesis that those true sentences that adults are quickest to verify are assessed correctly by children at the earliest age.

The present study was designed to test the hypothesis that children's error-rate differences would also parallel adult decision latencies for false sentences such as "Some fruits are vegetables." This type of sentence has been termed contradictory by Holyoak and Glass (1975), since the sub-

ject and predicate words are incompatible in meaning. Holyoak and Glass had adults generate such sentences by asking them to provide false completions for sentence frames such as "Some fruits are ———— ." The frequency with which different predicate words were produced was tabulated. These production-frequency (PF) norms were then used to select sentences for a verification experiment. The contradictory sentences were divided into three groups. High-PF sentences (e.g., "Some fruits are vegetables") were frequently produced by normative subjects and contained subject and predicate words with a common immediate superordinate. Low-PF sentences (e.g., "Some fruits are flowers") were produced much less frequently but still contained subject and predicate words with a close common superordinate. Finally, anomalous sentences (e.g., "Some fruits are hills") were produced infrequently and contained subject and predicate words that were related only at a very abstract semantic level.

These contradictory sentences were randomly interspersed with an equal number of true sentences and presented one at a time to adult subjects instructed to evaluate each sentence as quickly as possible. The results indicated that anomalous sentences were rejected most quickly, followed by high-PF sentences, with low-PF sentences requiring the most time to disconfirm (see Glass et al. 1974). While these results may be interpreted in different ways, one explanation is in terms of the different kinds of falsifying evidence available to disconfirm each type of sentence. The subject and predicate concepts in an anomalous sentence have many incompatible properties, so that a person is almost certain to find quickly some difference between the two categories sufficient to justify a "false" decision. The subject and predicate concepts in high- and low-PF sentences, in contrast, have far fewer incompatible properties. The person therefore has less information available to determine that the sentence is false. However, in high-PF sentences the subject and predicate concepts have the special property of forming explicit semantic contrasts—for example, "fruit" and "vegetable" are antonyms or opposites. The subjects and predicates of low-PF sentences do not form such clear contrastive pairs. A direct semantic contrast between the subject and predicate categories may function as a readily available piece of disconfirming information, allowing subjects to reject high-PF sentences relatively quickly.

The present study investigated the relative age at which children first learn to evaluate correctly anomalous, high-PF, and low-PF sentences.

We wished to discover whether children's error-rate differences parallel adult verification speed for contradictory sentences and, in particular, whether a direct semantic contrast can aid children as well as adults in determining that a sentence is false.

Method

A total of 69 children (34 boys and 35 girls, with 18 in grade 1, 23 in grade 3, and 28 in grade 5) were read sentences of the form "Some S are P," where S and P were concrete nouns. They were asked to decide whether each sentence was true or false and were allowed as much time as they needed for each decision. At the beginning of the session, the experimenter said, "Your name is [child's name]," and asked whether that was true or false. After the child answered, the experimenter said, "My name is [child's name]," and repeated the question. All the children appeared to understand readily the distinction between true and false. After the sentences had been presented, the child was asked to match the subject and predicate words in the false sentences against line-drawn pictures on 3 × 5-inch cards, in order to ascertain whether the child could identify referents of the words. Presentation order in both the verification and picture-matching tasks was random. The experimenter, who knew nothing of the hypothesis being investigated, tested the children individually. Each session took approximately 10 min, sometimes slightly longer for the first graders.

Eight triplets of contradictory false sentences were selected from the production-frequency norms of Holyoak and Glass (1975) and more recent norms collected in our laboratory by D. Billman. Each triplet consisted of an anomalous sentence, a high-PF contradictory sentence, and a low-PF contradiction. All these sentences are listed in table 1. The high- and low-PF contradictions differ from the anomalous sentences in that the former contain subject and predicate concepts that share a close superordinate, while the subjects and predicates of anomalous sentences are related only at a more abstract level. The high- and low-PF sentences differed in the frequency with which their predicates were produced by normative adult subjects. A mean of 45% of respondents produced the high-PF sentences, while only 5% of respondents produced the low-PF sentences. The subject and predicate words in the high-PF sentences are close contrastive associates. The semantic relations between the subject and predicate words in the low-PF sentences are more diverse, but in each case the two categories share a close superordinate but do not form a direct semantic contrast.

TABLE 1

CONTRADICTORY FALSE SENTENCES
USED IN STUDY

QUANTIFIED SUBJECT	PREDICATE WORDS		
	Anomalous	High-PF	Low-PF
Some women are ..	trees	men	babies
Some horses are ...	tables	cows	donkeys
Some kings are ...	chairs	queens	women
Some salt is	rain	pepper	sugar
Some boys are	trees	girls	monkeys
Some apples are ...	cars	oranges	vegetables
Some coffee is	butter	tea	milk
Some valleys are ..	boats	mountains	lakes

The particular sentences used were chosen because their subject and predicate words were picturable and likely to be familiar to first-grade children. It has been suggested that the process by which children evaluate sentences is influenced by perceptual properties of the referents of words (Kosslyn, Note 1). It might therefore be the case that children are more likely to respond incorrectly "true" to a contradictory sentence if the subject and predicate are relatively similar in appearance. We therefore obtained ratings from both adults and children of the perceptual similarity of each subject-predicate pair used in the false sentences. A rating of 7 indicated that the two objects appeared identical, while a rating of 1 indicated that the two objects were totally unlike in appearance. Twenty-five undergraduates responded by writing a number from 1 to 7 beside each pair on a questionnaire. Eight first-graders responded by circling a number on a line for each pair as the pair was read aloud. The number 1 was labeled "not alike at all," and 7 was labeled "identical." For the college students, the mean ratings for pairs from the high-PF, low-PF, and anomalous sentences were 4.89, 4.77, and 2.17, respectively, while the first-graders produced corresponding means of 3.12, 3.10, and 1.52. Both adults and children always rated the pairs from anomalous sentences least similar in appearance, while the mean ratings for pairs from high- and low-PF sentences did not differ significantly. On the basis of perceptual similarity, one would therefore predict a relatively low error rate for anomalous sentences, but no difference between high- and low-PF sentences.

The items were divided into two sets. One set contained three false triplets and three true triplets that used the same subject word (a total of 18 sentences). The second set contained five triplets of false sentences and five pairs of true sentences (a total of 25 sentences). The true sentences were included solely to prevent the children from using the strategy of always responding "false." Each child evaluated just one item set, and an approximately equal number of children received each set.

All contradictory sentences have the property of being false, whether quantified by "some" or "all." However, in general "all" statements can be disconfirmed by finding a single counterexample, even if the subject and predicate are not contradictory (e.g., "All kings are fathers"). Accordingly, all sentences in the present study were quantified by "some."

Results

Table 2 presents the proportions of correct "false" responses made to each of the three types of contradictory sentences by subjects from each grade. The pattern of results was essentially the same for both item sets, so the data from the two sets were collapsed prior to statistical analysis. No child made an error on the picture-identification test, indicating that all children were able to use all the words in the false sentences correctly in a referential situation. The proportion of correct responses for the "true" filler items was virtually constant across the three age levels (0.88 for grade 1, 0.89 for grades 3 and 5). The different proportions of "true" responses to ostensibly false sentences across grades are therefore not attributable to differential biases toward responding "true."

Because of ceiling effects in some cells of the design, a number of separate analyses of variance were performed on the data. Two parallel sets of analyses were performed. In one case, mean proportions correct were obtained for each subject, collapsing across the different items in a cell. Differences among the various conditions were then tested with subjects as a random effect. In the second case, means were obtained for each of the eight triplets of items (e.g., "Some kings are chairs/queens/women"), collapsing across subjects. These latter analyses, which treated items as

TABLE 2

PROPORTION OF CORRECT "FALSE" RESPONSES
TO CONTRADICTORY SENTENCES

GRADE	SENTENCE TYPE		
	Anomalous	High-PF	Low-PF
1	0.99	0.87	0.68
3	1.00	0.87	0.85
5	1.00	0.93	0.91

random effect, were designed to test whether the obtained differences would be expected to generalize to other anomalous and contradictory sentences (Clark 1973). Because of the relatively small number of items used, significance levels were generally lower in the item analyses than in the subject analyses. Since the data were proportions, arc sine transformations were applied prior to all analyses, as recommended by Winer (1971, p. 400).

As the data in table 2 indicate, the hypothesis was fully supported that the earliest age at which children would consistently reject the various kinds of false sentences would parallel adult latency differences. For each grade level, an analysis of variance was performed on the differences in proportion correct for the three types of false sentences. First graders correctly rejected anomalous sentences more often than meaningful sentences (high- and low-PF items combined). This difference was significant across subjects, $F(1,34) = 13.1$, $p < .01$, and also across items, $F(1,14) = 7.90$, $p < .025$. Within the meaningful sentences, the first graders rejected high-PF items more often than low-PF items. This latter effect was highly significant with subjects as a random effect, $F(1,34) = 12.0$, $p < .01$, and approached significance with items as a random effect, $F(1,14) = 3.53$, $p < .10$.

By grade 3 the difference in proportion correct between high- and low-PF sentences had essentially disappeared and did not approach significance in either the subject or the item analyses. But anomalous sentences were still rejected more often than meaningful false sentences, $F(1,44) = 9.66$, $p < .01$, in the subject analysis and $F(1,14) = 4.52$, $p < .06$, in the item analysis. This latter trend still remained in the fifth-grade data, although the difference was significant only across subjects, $F(1,54) = 14.2$, $p < .01$, and not across items, $F(1,14) = 2.93$, $p < .25$.

As can be seen from table 2, the increase in performance for each item type was monotonic across the three age levels. Separate analyses tested the age trend for each condition. The monotonic increase was highly significant for the low-PF items, both across subjects, $F(1,66) = 12.7$, $p < .01$, and across items, $F(1,14) = 16.2$, $p < .025$, by items. A monotonic trend was also present for the high-PF items, although the significance levels were reduced, $F(1,66) = 4.53$, $p < .05$, by subjects and $F(1,14) = 8.05$, $p < .025$, by items. In all these analyses, the quadratic component of the age trend was nonsignificant, $F < 1$. Performance on the anomalous sentences was at the ceiling for all three grades.

Discussion

The present results for false sentences, together with the results obtained for true sentences by Rosch (1973), suggest that the organization of semantic information in memory is similar for children and adults. There are at least two possible explanations for the parallels that emerged from the present study. First, it may be that those semantic relationships evaluated most easily by adults are learned at the earliest ages. For example, it may be that the first graders had not always acquired the information necessary to determine that a low-PF sentence is false. This is particularly plausible in that the decisions were made without any overt time pressure. Second, it could be that even the youngest children in the present study had available in memory the same semantic information as adults but that the younger children do not search for information as deeply or exhaustively as do older children and adults. This would explain why younger children make most errors on those items that adults evaluate most slowly. While the present results do not distinguish between these possibilities, they do suggest that the relative availability of semantic information fits a similar pattern for first-grade and older age groups.

Perhaps the most striking result was the fact that performance with the anomalous sentences was already virtually perfect by the first grade. This is consistent with the results of James and Miller (1973), who found that children as young as 4–8 can detect anomalous word combinations in sentences. More interesting, perhaps, was the lower error rate for high-PF than for low-PF sentences at the youngest age level. This difference, which parallels adult decision speed, was obtained even though the two groups of sentences did not differ in the perceptual similarity of the subjects and predicates. In addition, adult ratings indicate that the concepts in high-PF sentences are more similar in meaning than those in low-PF sentences (Holyoak & Glass 1975). The present results suggest that both children and adults can use a direct semantic contrast between the subject and predicate category (e.g., "fruit-vegetable") as evidence that the sentence is false.

It is worth noting that, even at the first-grade level, children were always able to pick out correctly a picture corresponding to all the words used in the false sentences. In some sense, they therefore knew what all the words meant and how to apply them to objects. But this ability did not guarantee that they would recognize the semantic relation—contradictoriness—that holds between certain word meanings in adult usage.

The present results are compatible with the view that the semantic knowledge people use to recognize contradictions in simple sentences develops by a continuous accumulation process. The hypothesis that adults and school-aged children use similar types of information (e.g., antonymy relations) in evaluating sentences appears to merit further investigation.

Reference Note

1. Kosslyn, S. M. Constructing visual images. Unpublished doctoral dissertation, Stanford University, 1974.

References

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: a replication and extension of the Connecticut Category norms. *Journal of Experimental Psychology Monograph*, 1969, **80** (3, Pt. 2).

Clark, H. H. The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 335–359.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 240–248.

Glass, A. L.; Holyoak, K. J.; & O'Dell, C. Production frequency and the verification of quantified statements. *Journal of Verbal Learning and Verbal Behavior*, 1974, **13**, 237–254.

Holyoak, K. J., & Glass, A. L. The role of contradictions and counter-examples in the rejection of false sentences. *Journal of Verbal Learning and Verbal Behavior*, 1975, **14**, 215–239.

James, S. L., & Miller, J. F. Children's awareness of semantic constraints in sentences. *Child Development*, 1973, **44**, 69–76.

Loftus, E. F. Category dominance, instance dominance, and categorization time. *Journal of Experimental Psychology*, 1973, **97**, 70–94.

Rips, L. J.; Shoben, E. J.; & Smith, E. E. Semantic distance and the vertification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 1–20.

Rosch, E. R. On the internal structure of perceptual and semantic categories. In T. M. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press, 1973.

Wilkins, A. T. Conjoint frequency, category size and categorization time. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 382–385.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971.