Semantic control of eye movements in picture scanning during sentence-picture verification

PATRICIA A. CARPENTER and MARCEL ADAM JUST*
Stanford University, Stanford, California 94305

Eye movements were monitored while people inspected a picture to determine whether it confirmed or disconfirmed a sentence they had just read. The sentences were quantified statements like *Few of the dots are red*, and the picture was a display of a large and a small subset of dots. It was found that the subset of dots a person fixated was determined by the semantic representation of the sentence rather than by its superficial referent. Since *Few of the dots are red* is semantically represented as a denial that the larger subset is red, people verifying the sentence fixate the larger subset, even though it is the small subset that is superficially referenced. By contrast, since *A minority of the dots are red* is semantically represented as an assertion that the smaller subset is red, people verifying this sentence fixate the smaller subset. A model of how people derive and compare representations of sentences and pictures is outlined.

This paper examines how the meaning of a sentence influences the eye movements of a person looking at a picture that the sentence describes. Of special interest are those sentences which are superficially similar but have different semantic representations. For example, *Few of the dots are red* and *A minority of the dots are red* superficially refer to a smaller subset of dots. However, previous research (Just & Carpenter, 1971) has shown differences in how the two sentences are internally represented and processed. *Few* is actually a negative—a denial that the larger subset of dots is red—and this information may be represented as [neg (larger subset be red)], where “neg” stands for the negation operator. On the other hand, *a minority* is an affirmation that the smaller subset is red, and this information may be represented by (smaller subset be red). The question posed was whether the eye movements would be directed by the surface structure of the sentences or by their underlying semantic representations.

The semantic representations for these sentences were inferred from reaction-time experiments using the principle that two representations are easier to compare (shorter reaction time) when they match than when they mismatch (Just & Carpenter, 1971). In that experiment, people were timed while they read sentences such as those above and then verified

them by looking at a picture that contained two subsets of dots, e.g., 14 red and 2 black dots. Such a picture could be unambiguously represented as either (larger subset be red) or as (smaller subset be black). For negative quantifiers like *Few of the dots are red*, the sentence was represented as [neg (larger subset be red)] and was compared to the representation of the larger subset in the picture: (larger subset be red) when false, (larger subset be black) when true. These representations were inferred from the result that false instances took less time to verify than true instances. This indicates that there was a match between sentence and picture representation for the false instances, and a mismatch for true instances, according to the congruence principle. For affirmative quantifiers like *A minority of the dots is red*, the sentence and picture were represented in terms of the smaller subset. This was inferred from the result that true instances took less time to verify than false instances in this case. These inferences were further supported in another experiment, in which people were told whether to represent the picture in terms of the larger or the smaller subset. Thus, the semantics of the quantifier determined whether people internally represented the sentence and picture in terms of the larger subset (as for *few*) or in terms of the smaller subset (as for *a minority*).

The present experiment examines whether the semantics of the quantifier determine which subset in the picture people will fixate. To do this, the large and small subsets were physically separated so that the locus of visual fixation could be determined. The hypothesis was that the sentence would control the locus of fixation; specifically, people would fixate the

same subset that occurred in the semantic representation of the sentence. The types of quantifiers investigated were: (1) syntactic negatives that reference the smaller subset but are represented as denials of a property of the larger subset, e.g., *few*; (2) affirmatives that reference and are represented in terms of the smaller subset, e.g., *a minority*; (3) affirmatives that reference and are represented in terms of the larger subset, e.g., *many*. Thus, the predictions were that for quantifiers like *few*, people would fixate the larger subset more often than the smaller one, whereas the opposite would occur with quantifiers like *a minority*. Finally, quantifiers like *many* should cause people to fixate the larger subset rather than the smaller subset.

METHOD

The experiment was a verification task in which the person read a sentence, and then, while his eye movements were recorded, looked at a picture to determine whether the sentence was true or false of the picture. Half of the stimulus sentences appear in Table 1. The other half was identical, except that the word *black* replaced the word *red*. The three categories of quantifiers will hereafter be designated by the exemplars *few*, *a minority*, and *many*. Every sentence was presented on two separate occasions, each time matched with a different one of two pictures. The two possible pictures contained two dots of one color (red or black) at the top and 14 dots of the other color at the bottom. For example, the picture that falsified the sentence *A minority of the dots are red* had 14 red dots and two black dots. The two dots at the top were 4 deg apart, while the larger subset was about 12 deg square. Each sentence subtended a visual angle of about 18 deg. Although the sentence presentation ended before the picture appeared, the sentence was located on the vertical axis about 6 deg below the two dots at the top and 6 deg above the uppermost dots in the large subset, as shown in Fig. 1. The sentences and pictures were projected through two slide projectors, with tachistoscopic shutters, onto a screen 35.6 cm from the O.

Each trial consisted of a verbal “ready” signal, then a 1,500-msec presentation of the sentence, followed by a 700-msec presentation of the picture. The picture onset occurred precisely when the sentence presentation ended. The participants signaled their responses through a two-button decision apparatus, whose right-hand button always designated “true.” They were told that it was a

*The order of authors was decided by the toss of a coin. This was a completely collaborative effort. After August 1972, reprint requests should be addressed to: Psychology Department, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213. We wish to thank Herbert H. Clark for his constructive comments and Norman Mackworth for his advice and for the use of his eye camera.

Table 1
Stimulus Sentences

Type of Quantifier and Its Representation	Exemplars
A minority (Smaller subset be red)	A minority of the dots are red. About 2 of the 16 dots are red. A small proportion of the dots are red.
Few [Neg (larger subset be red)]	Few of the dots are red. Scarcely any of the dots are red. Hardly any of the dots are red.
Many (Larger subset be red)	Many of the dots are red. Most of the dots are red. Lots of the dots are red. A majority of the dots are red. About 14 of the 16 dots are red. A large proportion of the dots are red.

Fig. 1. Schematic representation of the relative positions of the sentence and the two subsets in the picture.

Table 2

Total Number of Fixations on the Two Subsets for the Three Types of Quantifiers

Example of Quantifier	Subset Fixated			Errors (Percent)
	Small	Large	Neither	
A Minority	92	49	55	9
Few	56	77	66	8
Many	27	253	135	4

Table 3

Number of Fixations on the Two Subsets as a Function of
Whether the Problem was True or False

Example of Quantifier	True/False	Subset Fixated		t Test	p
		Small	Large		
A Minority	True	53	19	4.14	.001
	False	39	30		
Few	True	36	33	2.76	.02
	False	20	44		
Many	True	5	140	5.17	.001
	False	22	113		

reaction-time experiment and, therefore, to answer as quickly as they could. There were 5 practice trials, as well as 48 test trials, in one random order for half the people and in another random order for the other half. From the instructions and the practice trials, the participants always knew where the large and small subsets would be in the picture. The 18 paid

participants were each tested for 20 min.

Eye movements during the picture scanning were recorded on videotape by means of a wide-angle reflection eye camera (Mackworth, 1968). This camera photographs the right eye and superimposes on it an array of fluorescent lights which frames the picture. When the eye moves, the array

remains in the same position, but the relative position of the pupil with respect to the array changes. Thus, fixation locus can be determined by noting which part of the frame of lights is centered in the pupil. The position of the light array on the pupil when the person read the sentence provided the baseline from which eye movements were scored. An eye movement from the baseline to either subset changed the relative position of the pupil by about 3.8 cm on the monitor, making the movements fairly distinguishable. For each trial, the person was assigned a single score based on the first vertical eye movement during the presentation of the picture. That score had one of three possible values: fixation on the small subset, fixation on the large subset, or no apparent eye movement from the baseline. This third alternative occurred because it was possible for people not to fixate either subset and still respond correctly. The scoring was done by one E who did not know which sentences had been presented.

RESULTS

The analysis of the eye-movement data included only those trials on which a correct true-false response was made and was collapsed over the adjectives *red* and *black*. All of the exemplars of the same type had similar results so the exemplars were treated as replications. For each O, there were 12 trials with a quantifier like *a minority*, 12 with *few*, and 24 with *many*.

The results, shown in Table 2, confirm the hypothesis that quantifiers like *a minority* cause fixation primarily on the small subset, while quantifiers like *few* cause fixation primarily on the large subset. The interaction between the type of quantifier and the locus of fixation was computed for each person by contrasting the difference between the number of fixations on the small and large subsets for *a minority* with the same difference for *few*. The interaction was significant, $t(17) = 4.82$, $p < .001$, and highly consistent across people. The interaction went in the wrong direction for only 1 out of the 18 participants, and even then there was a minimal difference of one fixation. The second part of the hypothesis was that quantifiers that reference the large subset, e.g., *many*, would direct fixation to the large subset. This prediction was supported by the greater number of fixations on the large subset than on the small subset [$t(17) = 7.02$, $p < .001$]. This difference was in the right direction for every person.

An unexpected finding was a significant difference in fixation locus

between true and false instances for each quantifier type. Since the truth value of the problem cannot be determined until after comparison of the sentence and picture, it was not expected that truth value would have an effect on which subset the person fixated. As shown in Table 3, however, a difference occurs with each type of quantifier as a function of whether the problem was true or false. For quantifiers that reference the small subset, i.e., *a minority* and *few*, there is a greater tendency to fixate the small subset when the sentence is true than when it is false. For quantifiers that reference the large subset, i.e., *many*, there is a greater tendency to fixate the small subset when the sentence is false than when it is true. The interaction of fixation locus and true-false was significant for each type of quantifier, as the *t* tests show in Table 3. These results, which will be discussed below, may be due to people sometimes choosing which subset of the picture to fixate on the basis of the color mentioned in the sentence.

DISCUSSION

This study shows how the representation of semantic content can influence how visual input will be inspected and interpreted. Superficially similar quantifiers with different semantic representations cause pictures to be scanned and represented in terms of different quantities in the picture. This illustrates how the relationship between the surface structure of a stimulus and its representation or interpretation is flexible. Sentences can affect the interpretation of pictures, and vice versa. Different sentences that specify "what's in" a picture may result in people using different scanning strategies or search images to code the picture. For example, Clark (in press) reported an experiment in which he presented people with a circle on a white background and described it as "a hole." When the circle was gone, the paper was entirely white and people were told either that the *hole is absent* or that a *lid is present*. Clark used reaction times to show that these two sentences, applied to the same visual stimulus, caused two different representations of the picture. Other interesting examples of the flexibility of the stimulus-representation relation are ambiguous sentences and figures. Both are stimuli with more than one likely representation, and the interesting question is how various factors give one representation preference over the others.

A simple model can account for how the sentence and picture representations interact in this task.

The task requires that the person first read and represent the sentence, look at and represent the picture, and then compare the two representations. For example, when the person reads *Few of the dots are red*, the information in the sentence is represented as a denial that many of the dots are red: [neg (larger subset be red)], where "neg" stands for the negation operator. After representing the sentence, he fixates the subset of the picture that is specified in the sentence representation—in this case, the larger subset—and represents the picture in terms of this subset. Thus, for the true case, the person would fixate on the subset of 14 black dots and code the picture as (larger subset be black). This information from the picture is then compared with the sentence representation. In the comparison process, the color predicates of the sentence and the picture representations mismatch (red vs black). A second mismatch occurs because the sentence representation contains a negation operator, while the picture representation doesn't. The person has a rule that assigns the response *true* to even numbers of mismatches and *false* to an odd number of mismatches, and then executes the appropriate response, which in this case is *true*.

The processing model outlined above cannot account for the fixation differences that were found between true and false problems. The model says that in order to determine truth value, people fixate the subset specified by the quantifier in the sentence representation. And overall, the quantifier did have a dominant influence over the choice of subset to fixate. However, the true-false difference can be explained if one postulates that on a small percentage of trials, Os fixated the subset specified by the color term in the sentence. To do this, the O would first have to determine the location of the color with his peripheral vision, because he never knew if the red subset, for example, would be on the top or bottom. This strategy of occasionally matching on the basis of color interacts with the main hypothesis to produce the obtained true-false differences. For example, the main hypothesis assumes that for a quantifier like *a minority*, the person would fixate the small subset. In true instances, the color in the sentence predicate and the color of the small subset would match, so both strategies would cause the person to code the small subset. But, for false instances, the color in the sentence and the color in the small subset of the picture do not agree. So, if a person fixated on the basis of the color mentioned in the

sentence, he would fixate the large subset. This analysis predicts more fixations on the small subset for true than for false problems and, conversely, more fixations on the large subset for false than for true problems. This is, in fact, the case (see Table 3). This reasoning predicts the same trend for *few* and the opposite for *many*—all of which are found in the data. Although this explanation of peripheral color detection is post hoc, it is consistent with Williams's (1966) finding that color is a strong factor in fixation-target selection. With a filming technique that allows finer temporal resolution, one could determine if the eye movements said to be influenced by color matching occur later in the 700-msec interval than do those which are primarily influenced by the quantifier. If the discrimination specified by the sentence predicate had been more complex, like a shape discrimination, then it probably could not have been performed as easily peripherally. For example, a task with displays of squares and circles and sentences like *Many of the figures are square* might have prevented Ss from using peripheral vision in selecting the fixation target. As it was, Ss may have sometimes determined the color peripherally. This may also account for why Ss often made correct true-false responses without making an eye movement detectable with our technique.

There are asymmetries between the two types of affirmative quantifiers, *a minority* and *many*, that have not been mentioned. Although both quantifiers seem to have representations that mirror their surface structure, *many* caused fixations on the large subset more often than *a minority* caused fixations on the small subset. There may be a preference for fixating and representing the picture in terms of the general condition (the large subset) rather than the exceptions. However, since the large subset was always on the bottom of the picture, subset size was confounded with position. This result must therefore be evaluated with caution.

This experiment explores part of the general question of how people internally represent sentences and pictures in a verification task. The structure of an internal representation of a sentence can determine how a person will subsequently scan a picture. Furthermore, the scanning pattern is correlated with the structure of the internal representation of that picture, as inferred from reaction-times studies (Just & Carpenter, 1971). Thus, people scan and code pictures in a way that is

congruent with the representation of linguistic information. It now remains to specify how other interesting semantic structures besides quantifiers can direct the way people code perceptual events.

REFERENCES

CLARK, H. H. Semantics and

comprehension. In T. A. Sebeok (Ed.), *Current trends in linguistics*. Vol. 12. *Linguistics and adjacent arts and sciences*. The Hague: Mouton, in press.
JUST, M. A., & CARPENTER, P. A. Comprehension of negation with quantification. *Journal of Verbal Learning & Verbal Behavior*, 1971, 10, 244-253.

MACKWORTH, N. H. The wide-angle

reflection eye camera for visual choice and pupil size. *Perception & Psychophysics*, 1968, 3, 32-34.
WILLIAMS, L. G. The effect of target specification on objects fixated during visual search. *Perception & Psychophysics*, 1966, 1, 315-318.

(Accepted for publication March 6, 1972.)