Semantic congruity and lexical marking in symbolic comparisons: An expectancy hypothesis

MARC MARSCHARK and ALLAN PAIVIO
University of Western Ontario, London, Ontario N6A 5C2, Canada

Lexical marking and semantic congruity effects were investigated in four symbolic size comparison experiments. Predictions followed from an expectancy hypothesis suggested by results of recent comparative judgment studies. According to the present position, lexical marking and semantic congruity should be mutually exclusive effects in such tasks, and the demonstration of either dependent upon the order in which the stimuli and the comparative term are evaluated. When the comparative precedes the stimuli, an expectancy is created whereby the subject is more likely to be prepared for, say, large items following the comparative "larger" and small items following the comparative "smaller." In addition, the usual advantage of unmarked as compared to marked comparisons should be offset by the initial processing of the comparative. As predicted, the comparative-stimulus presentation order produced a significant semantic congruity effect and no effect of lexical marking in Experiment 1. Conversely, when stimuli precede the comparative, or are presented simultaneously with it, no expectancy should be created, as the items are immediately available to the subject, and the semantic congruity effect should not be obtained. Upon presentation of the comparative, however, unmarked comparisons should be easier than marked comparisons. Experiments 2 and 4 confirmed these expectations, as significant lexical marking effects were obtained and significant congruity effects were not. These findings are contrary to predictions derived from a semantic coding interpretation of the symbolic comparison process.

A number of recent investigations concerning the nature of symbolic information in long-term memory have made use of tasks requiring comparative judgments on various attribute dimensions. The typical experiment involves the presentation of pairs of stimuli (such as names or pictures of familiar objects or animals) and the measurement of subjects' reaction times to decide which of the two is, for example, larger in real life. Any comprehensive model of these symbolic comparisons must be able to account for several findings that are quite robust across a variety of similar tasks. Moyer (1973), for example, demonstrated that the distribution of reaction times for such judgments, given animal names as stimuli, follows a regular inverse function (similar to that obtained for perceptual judgments) according to the actual size differences between the two animals. Paivio (1975) extended this finding to decisions across as well as within conceptual categories. Further, he obtained a similar "symbolic distance effect" but consistently faster reaction times with pictures, as compared to words, as the stimuli.

The symbolic distance effect is usually accompanied by a "semantic congruity effect" when bipolar compara-

tives such as *larger/smaller* are presented in mixed-list trials. The congruity effect is essentially an interaction between the magnitude range of the stimuli and the comparative on which the judgment is made. When judgments on one comparative are faster overall than judgments on the other, the effect (when graphed) appears funnel shaped and hence is referred to as a "funnel effect." Otherwise, the functions simply cross each other, yielding a "crossover effect." In one demonstration of the latter, Holyoak and Walker (1976) showed that it is "easier to choose the *longer* term of the pair *decade/century* and the *shorter* term of the pair *second/minute*" (p. 287). The present study evaluates two possible accounts of the semantic congruity effect: a semantic coding model (Banks, 1977; Banks, Clark, & Lucy, 1975; Banks & Flora, 1977; Banks, Fujii, & Kayra-Stuart, 1976) and an expectancy hypothesis.

Banks' (1977) two-stage semantic coding model attributes the semantic congruity effect in symbolic comparison tasks to a mismatch between semantic codes assigned to the stimuli and the comparative on which they are to be judged. For example, if two stimuli are initially encoded as being relatively large, they will receive the binary codes *LARGE* and *LARGE+*. When the comparative "larger" is presented, it is encoded as *LARGE+* and thus, by a simple match process, the larger member of the stimulus pair is selected. In the event that "smaller" is presented (and coded as

This research was supported by Grant A0087 to the second author from the National Research Council of Canada. A version of this paper was presented to the annual meeting of the Canadian Psychological Association, June 1977.

SMALL+), the codes do not match and must be further transformed. The transformation is assumed to take some time, however, thereby producing the congruity effect. According to Banks and his colleagues (Banks et al., 1975; Banks et al., 1976), these code mismatches are a consequence of the independence of a subjects' perceptual codes for the stimuli from the code assigned to the comparative on which they are to be judged. Another consequence of this independence is that the model is explicitly insensitive to different (comparative/stimulus) presentation orders and predicts congruity effects regardless of which is presented first (Banks & Flora, 1977).

The expectancy hypothesis most clearly differs from the semantic coding model in the assumption that the semantic congruity effect is a consequence of dependence rather than independence in the processing of stimuli and comparative. The congruity effect is assumed to depend simply upon the fact that, in the usual symbolic comparison task, the comparative is presented prior to the stimulus pair, creating a situation wherein the subject may prepare in some way for the range of stimuli (e.g., small or large items) that will be presented. In some cases, this "expectation" will prove correct, thereby reducing reaction times, and in others incorrect, increasing reaction times. Over several trials, a congruity effect is obtained. Depending on one's theoretical preference, this preparation may be viewed as the establishing of a threshold (Banks et al., 1975), sampling possible size values (Moyer & Bayer, 1976; Anderson, Note 1), associative priming by the verbal "cue" of the comparative, or as a simple expectation or response bias for particular stimulus items. The latter possibility is especially likely in a finite stimulus set paradigm where the entire stimulus set may be known. Regardless of the nature of the hypothesized preparation state, the result is that the comparative creates an expectancy that would not be induced if the presentation order were reversed so that the subjects viewed the stimuli prior to the comparative or viewed both simultaneously.

Both the semantic coding model and the expectancy hypothesis can account for the congruity effects in all of the studies where it has been reported and, at first glance, there may seem to be no reason to prefer one model over the other. The semantic coding model has the advantage of being more formal than the expectancy hypothesis (but, see Moyer & Bayer, 1976, for a relatively formal expectancy model), but the latter does have two other advantages. First, the semantic coding model virtually always predicts congruity effects in bipolar-adjective symbolic comparison tasks, whereas the expectancy hypothesis predicts that under certain circumstances no effect should be obtained. The expectancy hypothesis is thus more amenable to evaluation than the semantic coding model and better fits the available data. A related problem with the semantic

coding model arises from the assumption that picture stimuli must be transformed from analog to digital codes prior to the critical comparison/decision process (Banks et al., 1975, p. 36). In addition to appearing rather unparsimonious, this assumption begs the question of the basis on which the actual code generation is accomplished so that one item is coded as larger (L+) or smaller (S+) than the other. By judicious selection of coding thresholds, this untestable aspect of the model allows it to predict virtually any form of the congruity effect in almost any experimental situation.

A second problem with the semantic coding model is that it has no a priori mechanism that would account for the lexical marking effects often found in symbolic comparison experiments. The explanation of the funnel effect supplied by Banks et al. (1975), however, does indicate a way in which the model could account for marking effects. They suggested that if there is a bias toward coding the stimuli in a particular way (e.g., a preference for coding items as large, resulting in a main effect of stimulus magnitude), all choices for that direction would be faster than the other, although a congruity effect would still be obtained. This implies the testable prediction that whenever there is such an encoding bias, the lexical marking effect should be obtained. The strength of the expectancy hypothesis, in this regard, is that, although it does not incorporate lexical marking *per se*, the assumptions on which it is based do suggest explicit predictions concerning the presence or absence of marking effects in different experimental situations. These assumptions further lead to the corollary that marking and congruity should be mutually exclusive effects, the demonstration of either depending on whether the comparative or stimulus pair is presented first. Essentially, this claim is a consequence of the fact that the conditions under which expectancies are created (and congruity effects are obtained) are incompatible with those necessary for the demonstration of a lexical marking effect. When the comparative is presented first, it remains visible (in most experiments) for 1-3 sec before the stimuli are presented. This time should be sufficient for "complete" processing of either member of a bipolar-adjective pair, thus offsetting the usual advantage of the unmarked member.¹ In this case, the subjects' expectations concerning possible stimuli (consistent with the preceding comparison term) would result in the observed semantic congruity effect. In contrast, when the stimuli are presented first, no expectations should be created (since the stimuli are immediately available), and a congruity effect would not be expected. Rather, since a response is required immediately after the comparative is presented, the usual effect of lexical marking would be expected, whereby faster responses are obtained with unmarked as compared to marked comparison terms.

Our claim that semantic congruity and lexical

marking effects are mutually exclusive will probably go against the grain of prevailing opinion in this area. However, the original formulation of the expectancy hypothesis derived solely from our observation of a pattern of results in the relevant literature indicating consistent support for this claim. It is important to note, in this regard, that we restrict our discussion to symbolic comparisons. With one possible exception (*viz.*, Banks & Flora, 1977), which we will examine presently, the symbolic comparison literature is consistently in accord with predictions from the expectancy hypothesis and its corollary. That is, those studies in which the comparative has preceded the stimulus pair in a mixed, bipolar-adjective judgment task have demonstrated significant semantic congruity effects but have not obtained the typically robust lexical marking effect (e.g., Banks, 1977; Banks et al., 1975; Banks & Flora, 1977, Experiment 2;² Holyoak & Walker, 1976). Conversely, those studies in which the comparative has followed or has been presented simultaneously with the stimulus pair have obtained only the lexical marking effect (e.g., Moyer & Landauer, 1967; see also, Glushko & Cooper, 1978, Experiment 2).

Although the expectancy hypothesis will be considered here only in terms of symbolic comparison (*i.e.*, long-term memory) tasks, previous studies concerned with the semantic congruity effect have sometimes cited perceptual comparison and other studies as demonstrating its co-occurrence with lexical marking in producing funnel effects. Unfortunately, while the congruity effect is quite robust in symbolic comparison tasks, it is somewhat variable in perceptual tasks. This inconsistency may be due to the greater variability in suitable strategies or simply to the more frequent use in perceptual studies of trials blocked by comparative rather than randomly intermixed. In blocked presentations, subjects may quickly discover that a bias in favor of the comparative (e.g., expecting large items on all "larger" trials) is an ineffective strategy on 50% of the trials. Alteration of the strategy to resist such expectancies (or sampling or threshold establishment) would then reduce or eliminate the congruity effect. In any case, this variability creates problems for any model that attempts a comprehensive explanation of the comparative judgment process. In terms of the semantic coding model, for example, Banks (1977, p. 111) suggests that a congruity effect will be obtained in such tasks only "if the *S* codes the stimuli in such a way that they are in some conditions congruent with one of the instructions and in others, congruent with the other. . . ." Similarly, the expectancy hypothesis would predict congruity only in situations where correct and incorrect expectancies are established (*cf.* MacLeod, Hunt, & Mathews, 1978). The fine distinction between the models, in this regard, may prove difficult to test, and in the meantime the relative merits of the two will remain indeterminate. Nevertheless, a brief discussion

of several relevant studies seems worthwhile, as there appears to have been some confusion concerning their findings.

The most frequently mentioned of the perceptual comparison studies were presented by Audley and Wallis (1964) and Wallis and Audley (1964), who were the first to extensively investigate the semantic congruity effect. Their research has been somewhat misinterpreted, however, as providing empirical demonstration of consistent funnel effects (e.g., Banks et al., 1975). Throughout this discussion, it should be remembered that the term "funnel effect" is typically used to describe a situation in which significant effects are obtained for both congruity and marking. Although Wallis and Audley did not specifically address the possibility of lexical marking effects in either paper, it is possible to assess this on the basis of data presented in at least some of their experiments. In fact, of their six relevant experiments, only two showed funnel-type congruity effects. Wallis and Audley (1964) conducted two experiments in which judgments of pitch were made in blocks of 12 trials each. One comparative ("higher" or "lower") was presented prior to each block and all trials within a block were made on that comparative, thus yielding essentially a comparative-stimulus paradigm. Here, then, the expectancy hypothesis would predict the demonstration of semantic congruity effects (as would the semantic coding model), and its corollary would preclude the demonstration of lexical marking effects. Examination of their presented data reveals that in Experiment A1, a typical lexical marking effect was not obtained and, in fact, was reversed (reaction time means = 716.8 for "lower" vs. 798.4 for "higher"). Experiment A2 also failed to obtain a lexical marking effect (reaction time means = 646.5 for "lower" vs. 647.0 for "higher"). Both experiments yielded apparent congruity effects. Audley and Wallis (1964) also used a blocked comparative-stimulus paradigm. Although they do not present enough data to assess adequately the possibility of marking effects, examination of their figures suggests that a marking effect was likely obtained in Experiments 1, 2, and 3, whereas it was not obtained in Experiment 4. Congruity effects were obtained in Experiments 1, 2 (funnel type), 3, and 4 (crossover type) (but, see Marschark & Paivio, Note 2).

Banks et al. (1975) also used a comparative-stimulus presentation order in two experiments involving judgments of height and length. As would be predicted from both the expectancy hypothesis and the semantic coding model, significant congruity effects were obtained in all cases. Consistent with the corollary of the expectancy hypothesis, the effect of lexical marking appears to have been negligible, despite the fact that there were biases (*i.e.*, faster response times) in both experiments for stimulus items coded as having greater magnitude (e.g., balloons coded as "higher"). Table 1

and Figures 2 and 4 of the Banks et al. (1975) paper, in fact, indicate that the typical marking effect was often reversed. If the semantic coding model were to make any predictions in this regard, they would have to favor the demonstration of a marking effect.

In summary, although the expectancy hypothesis and its corollary were not originally intended to apply to perceptual comparisons, our examination of several relevant studies has revealed that the majority obtained either a significant lexical marking effect or a semantic congruity effect, but not both (see also, Friend, 1973; Trabasso, Riley, & Wilson, 1975). Most, but not all, of these findings are in the direction predicted by the expectancy hypothesis.

Banks and Flora (1977) reported several experiments designed to test an expectancy hypothesis. In the first of these experiments (Experiment 3), 12 subjects received 36 trials each with comparative-stimulus and stimulus-comparative presentation orders. A set of only five items (creating eight pairs) was used, however, and only the data from two of the pairs (*ant-bat* and *cow-elk*) were analyzed. Although this design was intended to maximize the congruity effect, the stimulus-comparative presentation order yielded an effect of only 22 msec, sharply attenuated from the 89-msec effect obtained with the comparative-stimulus order. Banks and Flora note only that the overall effect of 56 msec was statistically significant ($p < .05$) and that the two orders did not significantly differ from each other in the size of their effects. It is thus unclear whether the smaller of the two congruity effects is itself reliable. Experiment 4 made use of the same stimuli and procedure as Experiment 3, but only six subjects were tested. Once again, a significant overall congruity effect was obtained ($p < .05$), but analyses of the independent effects for the two order conditions were not reported. The comparative-stimulus and stimulus-comparative presentation orders yielded congruity effects of 82 and 30 msec, respectively. The latter effect was again relatively small, and it is unclear from the information presented whether it would be reliable. However, the smallest congruity effect reported to be significant in the literature that we have surveyed was 36 msec (Banks & Flora, 1977, Experiment 2).

The last experiment in the Banks and Flora (1977) study involved an infinite-set stimulus paradigm. A stimulus-comparative presentation order was employed in which different subjects either were or were not instructed to refrain from stimulus processing until after the comparative was presented ("refrain" and "uninstructed" conditions, respectively). From the expectancy hypothesis, of course, the uninstructed group would not be expected to demonstrate a congruity effect. According to either the expectancy hypothesis or the semantic coding model, however, the instructions to postpone stimulus processing would essentially convert the functional aspect of the task

to that of the typical comparative-stimulus situation, insofar as subjects could not take advantage of the available stimuli (expectancy hypothesis) nor assign perceptual codes (semantic coding model) until after the comparative had been coded. Thus, either model would predict a semantic congruity effect in the refrain condition. Banks and Flora again reported statistical tests for the overall congruity effect and not those for the independent encoding conditions. The size of the effect for the uninstructed group, however, was only 30 msec, much smaller than the 65-msec effect obtained in the refrain condition. Although we are unable to assess the possibility of lexical marking from their presented data, the overall effect appears to be quite small (see Banks & Flora, 1977, Figure 6), consistent with the corollary of the expectancy hypothesis.

The following four experiments were designed to compare the expectancy hypothesis with the semantic coding model as an explanation of the semantic congruity effect. They were also intended to test the corollary of the expectancy hypothesis that semantic congruity and lexical marking effects are mutually exclusive and dependent upon the presentation order of the to-be-compared stimuli and the comparative on which they are to be evaluated. The first two experiments included picture and word stimuli as a between-subjects variable, but there appears to be no basis from which to predict differential congruity effects with the two types of material either from the expectancy hypothesis or from any current model of the symbolic comparison process. Contrary to the argument of Banks and Flora (1977, p. 279), we do not see any reason why a dual-code model would predict a semantic congruity effect for words but not for pictures. The symbolic comparison process for words (at least in the case of real-world object comparisons) seems to differ from that for pictures only insofar as it requires an additional transformation from a verbal code to some form of analog code. This transformation would be expected possibly to influence only the mean reaction time of the resulting curves and should in no way affect their slopes (cf. Paivio, 1975).

EXPERIMENT 1

Experiment 1 was intended to be a replication of earlier studies that have employed a comparative-stimulus presentation order. It explicitly evaluated, however, the corollary of the expectancy hypothesis that semantic congruity and lexical marking effects are mutually exclusive. Both the semantic coding model and the expectancy hypothesis predict the demonstration of a congruity effect in this experiment. The corollary of the hypothesis further predicts that a lexical marking effect will not be obtained. If the semantic coding model makes any prediction with regard to lexical marking, it would have to be dependent upon

whether a main effect of stimulus magnitude were demonstrated. That is, if there is a bias toward encoding stimulus items as being, say, relatively large, lexical marking should become operative. If there is no bias, a marking effect would not be predicted. It should be emphasized, however, that Banks et al. (1975) state merely that the semantic coding model "could" account for the marking effect (in the funnel effects obtained by Wallis and Audley, 1964) in this manner; however, the model does not appear to lend itself to any other predictions in this regard.

Method

Stimulus materials. Sixty mixed pairs of animal and object names were drawn from Paivio's (1975) norms. Pairs were constructed so that the ratios of the rated sizes of the two members of each pair would be in one of three ranges: 1-1.25 (hereafter referred to as Ratio 1.25), 1.5-2.0 (Ratio 1.75), or 2-2.5 (Ratio 2.25). Thirty-two pairs were in the range of Ratio 1.25, 16 pairs were in the range of Ratio 1.75, and 12 pairs were in the range of Ratio 2.25. Further, half of the pairs in each range were relatively small items (e.g., *book-knife*: ratio 1.19), and half were relatively large items (e.g., *camel-cow*: ratio 1.09). No two items were paired together more than once, although approximately 1/6 of the 120 items were repeated in different pairs, yielding different size ratios.

Four sets of stimulus materials were constructed: Two sets consisted of pictures of the paired items, and the other two sets consisted of the names of the paired items. The second list in each condition was obtained by reversing the left-right positions of the items in all 60 pairs of the first list. The two pictures of each pair were of approximately the same area (30-40 square cm) and were placed on a white card such that their approximate center points were 10 cm apart horizontally. The two words of each pair were typed in IBM Letter Gothic capitals, and their midpoints were approximately 3 cm apart horizontally.

Experimental design. Sixteen subjects received picture stimuli and 16 received word stimuli. Eight people were assigned to each of the two lists in a condition. As each list of 60 pairs was to be presented with a mixed list of comparatives (30 each with "larger" and "smaller"), a further counterbalancing was obtained by reversing the comparative "larger" to "smaller" (and vice versa) for four of the eight people receiving each list. This yielded a total of four sublists within each material condition, each of which was presented to four subjects. Each subject in any given cell received the stimuli in a different one of four presentation orders. In summary, this was a 2 (condition, pictures vs. words) by 2 (comparative, "larger" vs. "smaller") by 2 (relative item size, "large" vs. "small" end of the item pool) by 3 (ratio) by 4 (sublist) design, with subjects nested within condition and sublist.

Procedure. Subjects were instructed that they would see a series of picture or word pairs, each preceded by the word "larger" or "smaller." If the word "larger" appeared, they were to decide which of the two subsequent items was larger in real life and press a button under that item. Similarly, if "smaller" appeared, they were to decide which of the two items was smaller and press a button under that item. The instructions stressed both speed and accuracy.

All stimuli were presented in a three-field mirror tachistoscope. Each trial began with a 2-sec exposure of a fixation point that served as a ready signal. This was followed by a 1-sec exposure of the comparative, which in turn was followed immediately by a 4-sec exposure of the stimulus pair. Subjects responded by pressing one of two buttons placed under the tachistoscope approximately .5 m apart. Subjects rested their hands on the buttons between responses to facilitate rapid responding. Responses and reaction times from the onset of the stimulus pair until a button was pressed were recorded by a PDP-12 computer in an adjoining room.

Each subject received 10 practice trials (5 "larger" and 5 "smaller" comparisons) followed by 60 test trials. The entire experiment lasted 25-30 min.

Subjects. The subjects were 32 introductory psychology students at the University of Western Ontario; each received credit toward a course requirement. Subjects were randomly assigned to each cell of the design.

Results

Errors accounted for 8.5% and 8.9% of the responses in the picture and word conditions, respectively. Mean reaction times for correct decisions were computed within each subject for each ratio by relative item size by comparative cell. Twelve means were thus obtained for each subject. These data were analyzed by a 2 by 2 by 2 by 3 by 4 analysis of variance in which subjects was the random factor. Significant main effects were obtained for ratio [$F(2,48) = 45.80, p < .001$], reflecting a symbolic distance effect (Figure 1), and for relative item size [$F(1,24) = 40.80, p < .001$], indicating that two large items were easier to compare than two small items (i.e., a bias effect). As expected, the Relative Item Size by Comparative interaction was significant overall [$F(1,24) = 27.53, p < .001$], yielding a congruity effect of 36.5 msec (Figure 2). The picture and word con-

Figure 1. Reaction times for size judgments over three size ratio ranges (Experiment 1).

Figure 2. Reaction times for "larger" and "smaller" comparisons given either relatively large or relatively small items in the comparative-stimulus paradigm (Experiment 1).

ditions, independently, also yielded significant congruity effects [$F(1,12) = 29.53$, $p < .001$, and $F(1,12) = 4.70$, $p < .05$, respectively]. Relative item size and ratio also interacted significantly [$F(2,48) = 5.24$, $p < .01$], so that the faster latencies with two large things, as compared to two small things, increased as size ratio increased.

As predicted from the expectancy hypothesis corollary, and contrary to the prediction made on the basis of the semantic coding model, the effect of comparative (lexical marking) was not significant either overall or for the independent picture and word conditions (all F s < 1). The picture-word difference, although not significant, was in the expected direction (pictures faster than words) and, as can be seen in Figure 1, this factor interacted with ratio [$F(2,48) = 6.28$, $p < .005$]. Matched t tests between conditions at each ratio revealed a significant advantage for pictures at Ratio 1.75 [$t(48) = 2.12$, $p < .05$].

EXPERIMENT 2

Experiment 2 was identical to Experiment 1 except that it involved a stimulus-comparative rather than a comparative-stimulus paradigm. The expectancy hypothesis and its corollary, in this case, predicts a reversal of the effects obtained in Experiment 1: Here, a lexical marking effect should be obtained and a semantic congruity effect should not. The simple prediction of the semantic coding model, in contrast, is that a congruity effect should be obtained (Banks & Flora, 1977). Once again, any prediction of marking effects from the model would depend on the presence or absence of a main effect of stimulus magnitude.

Method

Design and Procedure. The stimuli in Experiment 2 were the same as those in Experiment 1, and the experimental design was identical. The only procedural change was that a 3-sec presentation of each stimulus pair immediately preceded rather than followed the comparative.

Following the last trial, all subjects were given a questionnaire concerning possible strategies that may have been used in the experimental task. "Frequency of use" was rated on the following 5-point scale: 1, used infrequently or never; 2, used occasionally; 3, used with about half of the pairs; 4, used very frequently; and 5, used with all or almost all of the pairs. Four likely strategies were questioned: (1) various verbal devices (e.g., comparing the paired objects verbally with some standard, word association, or pronouncing the names of the objects); (2) imagery, "only after the word *larger* or *smaller* came on" (a delayed processing strategy as intentionally induced by Banks and Flora, 1977, Experiment 5, p. 286); (3) imagery, used to make an immediate size comparison and thus be prepared for either comparative; and (4) "other." In order to gain some insight into a lexical marking effect possibly produced by the third strategy, a fifth question asked subjects how often they felt better prepared for one comparative than for the other.

Subjects. Twenty-four subjects were drawn from the same population as in Experiment 1. All received credit toward a course requirement and were randomly assigned to all cells.

Results

Errors accounted for 9% of the responses in each presentation condition. Mean reaction times for correct decisions were analyzed exactly as in Experiment 1. Significant main effects were again obtained for ratio [$F(2,32) = 29.22$, $p < .001$] (Figure 3) and relative item size [$F(1,24) = 12.04$, $p < .005$]. As predicted by both models, there was also a significant main effect of comparative (lexical marking) [$F(1,16) = 16.17$, $p < .001$]. The picture and word conditions, independently, also yielded significant marking effects [$F(1,8) = 9.52$, $p < .025$, and $F(1,8) = 7.05$, $p < .05$]. Consistent with the expectancy hypothesis, the overall congruity effect (Relative Item Size by Comparative interaction) was not significant [$F(1,16) < 1$] (see Figure 4), nor were the comparable interactions for the independent picture and word conditions. The Relative Item Size by Comparative by Ratio interaction was also not significant, suggesting the absence of a congruity effect across all ratios. The sizes of the effect were 29, 33, and 44 msec at Ratios 1.25, 1.75, and 2.25, respectively. The size of the overall nonsignificant

Figure 3. Reaction times for size judgments over three size ratio ranges (Experiment 2).

Figure 4. Reaction times for "larger" and "smaller" comparisons given either relatively large or relatively small items in the stimulus-comparative paradigm (Experiment 2)

effect was 29 msec, comparable to that obtained in Experiments 3, 4, and 5 of the Banks and Flora (1977) study.

Since the stimuli were available to the subjects for 3 sec prior to the presentation of the comparative, a significant main effect of presentation condition (pictures vs. words) was not expected. Presumably, this interval is sufficient for the recoding of either one or both types of the material into some common format, such as images, prior to the comparative judgment (cf. Paivio, 1977; Posner, Boies, Eichelman, & Taylor, 1969; Seymour, 1975). In fact, the picture-word difference was not significant [$F(1,16) < 1$].

Analyses of the postexperimental questionnaire indicated that there were no significant differences between those subjects who received picture stimuli and those who received word stimuli [all $F_s(1,4) \leq 3.00$, $p_s > .15$]. Examination of the mean frequency of usage ratings for each strategy revealed that verbal strategies were used only infrequently or never (mean = 1.12, collapsed over both conditions), and the delayed imagery processing only occasionally (1.88). The most frequently employed strategy was the immediate imaginal comparison of the two stimuli and decision of which was larger and which smaller (4.50) (cf. Banks & Flora, 1977, p. 286). Only occasionally, however, did subjects feel that they were more prepared for one comparative than for the other (2.33). All subjects who scored this last question reported "larger" to have the relative advantage.

EXPERIMENT 3

Both the demonstration of a lexical marking effect and the absence of a semantic congruity effect in Experiment 2 were clearly consistent with predictions from the expectancy hypothesis and its corollary. These findings, however, might also be interpreted in terms of the semantic coding model as follows: Because subjects were able to decide in advance which stimulus member was larger and which smaller, the codes of L+ and S+, respectively, could already have been assigned when the comparative was presented (although it is unclear from the model why these codes need be assigned if the relative size decisions have previously been made in order to establish the codes). Presentation of either "larger" (coded L+) or "smaller" (coded S+) would thus result in an immediate match, and no congruity effect would occur.

Experiment 3 provides a more definitive test of the alternative models in a case where the two clearly make opposite predictions. The situation is one in which the comparative and stimulus pair are presented simultaneously. According to the expectancy hypothesis, this availability of complete information would obviate the establishment of correct or incorrect expectancies, and thus a semantic congruity effect should not be obtained.

However, if subjects tend to process the available comparative prior to the stimulus pair, a trend toward congruity (and away from marking) should be observed. If the stimulus pair were to be processed prior to the comparative, the reverse would be expected. In either of the extreme (and unlikely) cases where all subjects always process the comparative or stimuli first, a significant congruity or marking effect, respectively, should be obtained.

Predictions from the semantic coding model are quite different. Because the coding of the comparative and stimuli are independent, the model, as always, predicts that a congruity effect should be obtained regardless of whether subjects process the comparative "before" or "after" (Banks & Flora, 1977). One might wish to argue that simultaneous presentation could be a special case in which the codes for comparative and stimuli are not independent. However, this situation is actually a less extreme case for dependence than that in which the comparative is presented prior to the stimulus. As Banks et al. (1975) note, "in the semantic coding model as we have presented it, the subject forms his perceptual codes with a certain degree of independence from the instruction just presented to him. For example, *the subject cannot set himself to code the balloons as LOW/LOW+ even when he has just read the instruction 'Which balloon is lower?'* Such a strategy would have been optimal" (p. 44, italics added). This inflexibility is seen to be a consequence of the need for fast responding in a reaction time task where "the comparison stage seizes on the earliest codes available for the stimuli" (Banks et al., 1976, p. 437).

Method

Stimulus materials. Seventy-two mixed pairs of animal and object names were drawn from the stimuli of Experiment 1 and a subsequent, unpublished study by these authors. The characteristics of the pairs were consistent with those described in Experiment 1. The pairs were randomly divided into two sets of 36 with the restriction that the three ratio ranges (1.25, 1.75, and 2.25) contain 12 pairs each. Half of the pairs in each range consisted of relatively large items, whereas the remainder contained relatively small items. Within each ratio by item size cell, half of the pairs were designated for "larger" comparisons and half for "smaller" comparisons. The right-left positions of the correct answers were suitably counterbalanced. All of the pairs were typed in IBM Letter Gothic capitals in the center of white cards such that their center points were approximately 3 cm apart horizontally. In one of the lists, the appropriate comparative was typed in Letter Gothic small letters 1 cm above the stimulus words, centered between them. The remaining list had the comparatives 1 cm below the pair. The purpose of this construction was simply that we had no knowledge of possible effects created by single above or below presentation. Comparative position was therefore included as a factor in the design. Picture stimuli were not included in this experiment, as the previous two have demonstrated comparable results with word and picture materials.

Design and Procedure. Presentation order of the two stimulus lists was counterbalanced across subjects. Each set of 36 test trials was preceded by 5 relevant practice trials. Each tachistoscopic trial was preceded by a warning tone and a 1-sec exposure

of a fixation point. Immediately following, a stimulus card appeared and remained until a response was given. Computer timing started with the exposure of a stimulus card and terminated with a response. Both speed and accuracy were stressed.

After the last trial, all participants were given a questionnaire concerning possible strategies used in the experimental task. Frequency of use was rated on the same 5-point scale used in the previous experiment. Two questions concerned the use of verbal and imaginal strategies (as in Experiment 2). The remaining two questioned whether subjects first made a decision about which animal was larger (and which was smaller) and then looked to see what the "question" was, or whether they first looked at the comparison term and then decided which member of the pair was the correct answer. A fifth question asked to which type of card the subjects felt it easier to respond: the pair on the top and the comparative on the bottom, the reverse, or both about the same. The entire experiment lasted approximately 20 min.

Subjects. Eighteen subjects were drawn from the same population as the previous experiments. All participants received credit toward a course requirement and were randomly assigned to all cells.

Results

Errors accounted for approximately 7% of the responses in each comparative position cell. Mean correct reaction times were analyzed by a 2 (presentation order, comparative above then below or below then above) by 2 (comparative position) by 2 (comparative) by 2 (relative item size) by 3 (ratio) analysis of variance, in which subjects was a random factor nested in presentation order. The analysis indicated that responses were significantly faster (by a mean of 74 msec) on trials in which the comparative appeared above the stimulus pair [$F(1,16) = 5.56, p < .05$], and that "smaller" comparisons were significantly faster than "larger" comparisons [$F(1,16) = 7.30, p < .025$].³ A strong symbolic distance effect was also obtained [$F(2,32) = 17.74, p < .0001$]. Significant two-way interactions were obtained for Presentation Order by Position [$F(1,16) = 6.91, p < .025$] and Relative Item Size by Position [$F(1,16) = 4.96, p < .05$]. The former resulted from comparative-above responses taking 150 msec longer than comparative-below responses only for the below-above order. There was no difference with the above-below order. The latter interaction resulted from comparisons of smaller items yielding faster reaction times only when the comparative was above the stimulus pair. These and two significant three-way interactions involving ratio do not appear to qualify the effects of interest here and will therefore not be considered further. There were no other significant effects.

As predicted from the expectancy hypothesis, and contrary to the semantic coding model, a significant semantic congruity effect was not found, although a trend in that direction yielded an effect of 54 msec (see Table 1). In terms of the expectancy hypothesis, this trend and the lack of a typical lexical marking effect suggest that subjects tended to process the comparatives prior to the stimulus pairs. This was verified by the

Table 1
Mean Correct Reaction Times in Seconds (Experiment 3)

Comparative	Item Size		Ratio		
	Small	Large	1.25	1.75	2.25
Smaller	1.71	1.72	1.84	1.66	1.64
Larger	1.82	1.73	1.90	1.74	1.69

questionnaire data, which indicated a significant overall preference for comparative-stimulus as compared to a stimulus-comparative processing order [$t(34) = 4.63, p < .01$]. Subjects also showed a higher rated preference for imagery, as compared to verbal strategies [$t(34) = 2.93, p < .01$]. Ratings did not differ as a function of presentation order for any of the questions [all $t(16) < 1.50$]. Nine of the 18 subjects indicated that they found the comparative-above-pair trials easiest. Four preferred the reverse, and five indicated that they found no difference.

EXPERIMENT 4

The final experiment again involved a paradigm in which comparative and stimuli were presented simultaneously, but this time the information was in the form of sentences requiring true-false responses (e.g., A goat is larger/smaller than a dog).⁴ This design has two advantages for the present purposes. First, it allows a test of the generality of previous symbolic comparison findings (and models designed to account for them) in a somewhat different paradigm. It specifically allows a further test of the alternative predictions from the semantic coding model and the expectancy hypothesis in a task situation similar to that we tried to establish in the previous two experiments. Second, this form of presentation seems more likely to induce concurrent processing of the presented information or, at worst, a bias toward a stimulus-comparative order. Thus, according to the expectancy hypothesis and its corollary, only an effect of lexical marking should be obtained, to the exclusion of a semantic congruity effect, since no correct or incorrect expectancies should be created. The semantic coding model, in contrast, would again predict the presence of a congruity effect. Suppose, for example, a subject saw the sentence, "A squirrel is larger than an ant." Assuming a left-to-right processing scheme, "squirrel" would be coded as S(+) (by the usual coding mechanism), "larger" would be coded as L+, and "ant" would be coded as S+, thereby creating the requisite mismatch for a congruity effect. In the event that the subject ignored the sentence format and simply processed the stimuli and comparative (in either order), the model would still predict a congruity effect for the reasons outlined in Experiment 3.

Method

Stimulus materials. A total of 120 different pairs of animal and object names were drawn from Paivio's (1975) norms.

The rated size ratios for 64 of the pairs were in the range of Ratio 1.25, 32 were in the range of Ratio 1.75, and 24 were in the range of Ratio 2.25. Half of the pairs in each range were relatively large items and the other half were relatively small items. Left-right position of the correct answer and assignment to "larger" or "smaller" comparisons were suitably counter-balanced in all cells in order to create 60 true and 60 false sentences. Further, four different lists were constructed, each of which contained all 120 pairs. Each pair appeared in a different form in each list, as in the following example: "An iron is larger than a rooster. A rooster is larger than an iron. An iron is smaller than a rooster. A rooster is smaller than an iron." Two forms of each sentence were thus true, and two false.

All sentences were typed in IBM Letter Gothic capitals and centered in the middle of white cards (without punctuation) for tachistoscopic presentation.

Design and Procedure. Each of the four lists was presented to five different subjects, and each subject within a list received the pairs in a different order. The experiment was thus a 4 (list) by 2 (relative item size) by 2 (comparative) by 2 (response, true or false) design, in which subjects was nested within list. As will be described below, ratio was not included as a factor, but was assessed separately for each subject.

Subjects were informed that their task was to decide whether each in a series of sentences was true or false. To simplify the demands on the subject, each used his/her dominant hand to respond "true" and the nondominant hand to respond "false." A pilot study that counterbalanced response side with hand dominance found no significant effects in that regard. Subjects each received 5 practice trials followed by three blocks of 40 test trials each. A 1-min rest period between blocks was permitted if desired. In all other respects, the procedure was identical to that of Experiment 3. The entire experiment lasted 25-30 min.

Subjects. Twenty University of Western Ontario summer school students were obtained through sign-up sheets. Each subject was paid \$2 for participating.

Results

Errors accounted for 15% and 11% of true and false responses, respectively. Mean reaction times for correct responses were analyzed by the appropriate 4 by 2 by 2 by 2 analysis of variance, in which subjects was the random factor. Significant main effects were obtained for relative item size [$F(1,16) = 16.12$, $p < .001$], indicating a bias for large items, response [$F(1,16) = 9.00$, $p < .01$], with "true" responses being faster than "false" responses, and, as would be predicted from either semantic coding or the corollary of the expectancy hypothesis, comparative (lexical marking) [$F(1,16) = 78.64$, $p < .001$]. The only significant interaction was Relative Item Size by Response [$F(1,16) = 7.11$, $p < .025$]. In particular, the congruity effect (relative item size by comparative) was not significant [$F(1,16) = 1.73$]; the size of the overall effect was 33 msec (see Table 2). This finding, of course, is consistent with the expectancy hypothesis and contrary to the prediction of the semantic coding model.

The robustness of the typical symbolic distance effect was investigated by examining it separately for each subject. Correlations were computed between the size ratio associated with each pair and the measured reaction time. Only correct responses were analyzed. Of the 20 subjects, 16 demonstrated significant distance effects

Table 2
Mean Correct Reaction Times in Seconds (Experiment 4)

Response	Comparative	Item Size	
		Small	Large
True	Smaller	2.42	2.41
	Larger	2.18	2.11
False	Smaller	2.54	2.40
	Larger	2.38	2.16

(i.e., significant inverse correlations), 10 at $p < .005$, 2 at $p < .01$, and 4 at $p < .05$ ($86 < df < 113$). Significant correlations ranged from $-.184$ to $-.418$. (The pilot study yielded significant distance effects for 15 of 20 subjects, similarly distributed.)

SUMMARY AND CONCLUSIONS

The present experiments investigated lexical marking and semantic congruity effects in symbolic comparisons. All of the findings were consistent with a simple expectancy hypothesis and its immediate corollary. According to this interpretation, when the comparative precedes the stimulus pair in such a task, the former creates an expectancy whereby the subject is more likely to be prepared for, say, large items following the comparative "larger" and small items following the comparative "smaller." In Experiment 1, this expectancy produced a significant congruity effect, whereas initial processing of the comparative apparently excluded a lexical marking effect. In contrast, when the comparative follows the stimulus pair (Experiment 2) or is presented simultaneously with it (Experiments 3 and 4), no expectancies are created since the subject can process the magnitudes of the to-be-compared stimuli prior to, or in conjunction with, the comparative. Accordingly, in Experiment 2, no congruity effect occurred, whereas a lexical marking effect was obtained since subjects did not have an exposure during which they could fully process the comparative terms. A lexical marking effect was not obtained in Experiment 3, as subjects tended to process the comparative prior to the stimulus pair. This strategy also produced a trend toward a congruity effect. Experiment 4, which biased subjects toward a concurrent or stimulus-comparative processing strategy, produced a large marking effect and a negligible congruity effect.

Both the semantic coding model and the expectancy hypothesis have been shown able to account for the congruity effect in all instances where it has been obtained. The expectancy hypothesis, however, appears to be the more general model, insofar as it also encompasses situations where the effect is not obtained and the semantic coding model makes erroneous predictions. Further, the present findings have supported a corollary of the expectancy hypothesis concerning

the mutual exclusion of lexical marking and congruity effects as a function of the symbolic comparison paradigm employed. These findings are obviously inconsistent with the semantic coding model or any other interpretation of the symbolic comparison process that explicitly predicts a congruity effect regardless of the processing order of stimuli and comparative.

REFERENCE NOTE

1. Anderson, R. C. *Knowledge of variability and the contextual determination of meaning*. Unpublished manuscript, 1977.
2. Marschark, M., & Paivio, A. *Semantic congruity and lexical marking in perceptual comparisons: The expectancy hypothesis*. Manuscript in preparation, 1979.

REFERENCES

- AUDLEY, R., & WALLIS, C. Response instructions and the speed of relative judgment. I. Some experiments on brightness discrimination. *British Journal of Psychology*, 1964, **55**, 59-73.
- BANKS, W. Encoding and processing of symbolic information in comparative judgment. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 11). New York: Academic Press, 1977.
- BANKS, W., CLARK, H., & LUCY, P. The locus of the semantic congruity effect in comparative judgments. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, **1**, 35-47.
- BANKS, W., & FLORA, J. Semantic and perceptual processes in symbolic comparisons. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, **3**, 278-290.
- BANKS, W., FUJII, M., & KAYRA-STUART, F. Semantic congruity effects in comparative judgments of magnitudes of digits. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, **2**, 435-447.
- FRIEND, K. Perceptual encoding in comparative judgments of race. *Memory & Cognition*, 1973, **1**, 80-84.
- GLUSHKO, R., & COOPER, L. Spatial comprehension and comparison processes in verification tasks. *Cognitive Psychology*, 1978, **10**, 391-421.
- HOLYOAK, K., & WALKER, J. Subjective magnitude information in semantic orderings. *Journal of Verbal Learning and Verbal Behavior*, 1976, **15**, 287-299.
- MACLEOD, C., HUNT, E., & MATHEWS, N. Individual differences in the verification of sentence-picture relationships. *Journal of Verbal Learning and Verbal Behavior*, 1978, **17**, 493-508.
- MOYER, R. Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics*, 1973, **13**, 180-184.
- MOYER, R., & BAYER, R. Mental comparison and the symbolic distance effect. *Cognitive Psychology*, 1976, **8**, 228-246.
- MOYER, R., & LANDAUER, T. Time required for judgments of numerical inequality. *Nature*, 1967, **215**, 1519-1520.
- PAIVIO, A. Perceptual comparisons through the mind's eye. *Memory & Cognition*, 1975, **3**, 635-647.
- PAIVIO, A. The relationship between verbal and perceptual codes. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception: Perceptual processing* (Vol. 9). New York: Academic Press, 1977.
- POSNER, M., BOIES, S., EICHELMAN, W., & TAYLOR, R. Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monograph*, 1969, **79**(1, Pt. 2).
- SEYMOUR, P. Semantic equivalence of verbal and pictorial displays. In A. Kennedy & A. Wilkes (Eds.), *Studies in long term memory*. New York: Wiley, 1975.
- TRABASSO, T., RILEY, C., & WILSON, E. The representation of linear order and spatial strategies in reasoning: A developmental study. In R. Falmagne (Ed.), *Reasoning: Representation and process*. Hillsdale, N.J.: Erlbaum, 1975.
- WALLIS, C., & AUDLEY, R. Response instructions and the speed of relative judgments. II. Pitch discrimination. *British Journal of Psychology*, 1964, **55**, 133-142.

NOTES

1. Although a variety of studies have investigated the linguistic bases and empirical consequences of the marking effect, we have yet to arrive at a satisfactory psychological mechanism that would account for it. The one common conclusion of most researchers in the area is that marked terms are in some sense more complex psychologically than their unmarked counterparts. This conclusion seems to do little more than restate the empirical observation that response latencies immediately following marked terms are longer than after unmarked terms. It does suggest, however, that marked terms may simply take more time to "unpack" than unmarked terms and that given sufficient processing time, the marking effect should disappear.
2. Experiment 1 of the Banks and Flora (1977) study did reveal an apparent lexical marking effect (see their Figure 1). This was apparently an artifact, however, resulting from several poorly identified stimulus drawings for half of the "small near" pairs.
3. Although the reason for this reversal of the marking effect is unclear, it is not unusual (e.g., Banks et al., 1975; Wallis & Audley, 1964).
4. This experiment was actually the first conducted in this series.

(Received for publication February 1, 1978;
revision accepted February 15, 1979.)