*Journal of Experimental Psychology:
Human Perception and Performance*
1975, Vol. 104, No. 1, 35–47

The Locus of the Semantic Congruity Effect in Comparative Judgments

William P. Banks
Pomona College

Herbert H. Clark and Peter Lucy
Stanford University

This paper describes and tests a two-stage model for a "semantic congruity effect" in comparative judgments. This effect is illustrated in the present experiments: When subjects were asked to choose the higher or the lower of two balloons tethered at the ends of strings, they were faster at choosing the higher of the two. But when asked to choose the higher or the lower of two yo-yos hanging at the ends of strings, they were faster at choosing the lower one. By hypothesis, this occurred because the balloons were coded at a first perceptual stage in terms of highness, and the yo-yos, in terms of lowness; then, at the second linguistic stage, the perceptual codes that matched the instructional codes ("choose the higher" or "the lower") resulted in the faster judgments. The present two experiments demonstrated that (a) the two stages are sequential, since changes in pairwise stimulus discriminability and in instructions had additive effects on the total reaction time, and (b) the presence of the semantic congruity effect depended on the actual perceptual codes applied to the stimuli.

Over the past 25 years investigators of comparative judgments have turned up a rather remarkable phenomenon we will call the *congruity effect.* The phenomenon is nicely illustrated in an experiment by Audley and Wallis (1964). Their subjects were shown either a pair of unequally intense but very bright lights, or a pair of unequally intense but very dim lights, all against a background of middle gray. The subjects, while timed, were required to pick out sometimes the brighter, and other times the dimmer, member of the presented pair. Because the two lights of each pair were quite discriminable, the subjects were almost always correct. Yet their reaction times revealed a curious interaction. When shown the pair of bright lights, the subjects were faster at picking out the brighter of the two. But when shown the pair of dim lights, they were faster at picking out the dimmer of the two. That is, the comparative judgments were faster when the direction of the judgment (brighter vs. dimmer) was in agreement with the relative level of the two stimuli on their underlying continuum (bright vs. dim). Hence the name "congruity effect." Since the two pairs of lights were identical for both the brighter and the dimmer choices, this effect must have resulted from semantic, or instructional, factors rather than from sensory factors alone.

Audley and Wallis described two varieties of the congruity effect, one they called the *crossover* effect and the other they called the *funnel* effect. The crossover effect is illustrated by the experiment just described (their Experiment IV). When the reaction times (RTs) of the judgments of "Which is brighter?" and "Which is dimmer?", to be called the brighter and dimmer choices, respectively, were plotted against the absolute intensity of the pair of lights judged, the lines actually "crossed over" in

---

The research reported here was supported in part by Public Health Service Grant MH-20021 from the National Institute of Mental Health. We thank Hiram H. Brownell for running Experiment 2, and John Taplin, Bonita Ford, Leah Light, and Roger Schmaltz for their helpful editorial comments on the paper.

Requests for reprints should be sent to William P. Banks, Department of Psychology, Pomona College, Claremont, California 91711.

a true interaction: The brighter choice was below the dimmer choice at high intensities, but the dimmer choice was below the brighter choice at low intensities. In another experiment, however, Audley and Wallis got a funnel effect. In this experiment, they used the same two pairs of lights, but presented them against a virtually black background. In this case, the brighter choices were faster for both pairs, but the difference in RTs was larger for the bright pair than for the dim pair. When plotted as before, the brighter and dimmer choices formed a funnel that widened at high intensities, hence the name *funnel effect.* Obviously, the funnel effect is also a type of congruity effect in that the comparative jugments are relatively faster when the direction of judgment is the same as the absolute level of the stimuli.

The congruity effect has been observed for a variety of perceptual and conceptual continua. The phenomenon was first reported by Shipley and his colleagues (Shipley, Coffin, & Hadsell, 1945; Shipley, Norris, & Roberts, 1946), who found a crossover effect in reaction times for color preferences. Wallis and Audley (1964) found funnel and crossover effects in judgments of pitch, and of brightness (Audley & Wallis, 1964). Ellis (1972) showed a crossover effect in comparative judgments of age; Marks (1972) showed a variety of crossover and funnel effects in judgments of the probabilities of events; and Friend (1973) showed a funnel effect in judgments of racial identity as defined by skin color. In addition, Clark, Carpenter, and Just (1973) showed something akin to both crossover and funnel effects in judgments of height, depth, size, and width.

There have been several attempts to account for these congruity effects. Audley and Wallis (1964) proposed a stochastic choice model, Wallis and Audley (1964) proposed a response competition model, and Marks (1972) proposed a discriminal dispersion model. In the present paper, we will propose yet another model, the semantic congruity model. The purpose of the experiments reported below was to evaluate these four models as accounts of the congruity effect. By investigating variables not previously studied, we intended to provide tests that the semantic congruity model would be able to pass, but the others would not. We will present the semantic congruity model here and save discussion of the other three models for the final section of the paper.

The semantic congruity model assumes that the mental operations intervening between the presentation of the stimuli and the making of the response can be broken down into a series of successive stages. There are two stages of particular interest, a first "perceptual" stage that generates a semantically coded description of the stimuli, and a second "linguistic" stage that makes use of these codes to determine the correct response. Further stages, such as response execution (cf. Smith, 1968), are assumed to wait for the outcome of the earlier stages before beginning to work and hence provide only additive components to the RT. Such later stages are not studied in the present experiments.

The perceptual stage can be conceived of as an "analogue-to-digital" processor. It takes the stimulus information and generates a coded semantic description for it. Although this stage undoubtedly consists of many substages, some of which may work in parallel and will not be additive, the important assumption is that it operates completely prior to the second, linguistic stage. Psychophysical variations in the stimuli—for example, variations in the absolute intensity and discriminability of the pair—should affect the speed and output of the perceptual stage, but have no direct effect on the linguistic stage. On the other hand, the linguistic stage takes the codes generated by the perceptual stage, matches them with the previously stored and coded instructions, and computes the correct response. This stage proceeds, we assume, in accordance with Clark's (1969) principle of congruence. By that principle it must find a match between the perceptual code and the linguistic code derived from the instruction before it can compute the correct response; when there is no match, it must operate on one of the codes to form a match, and this operation takes time to carry out.

According to the present model, the congruity effect is a consequence of operations in the linguistic stage. To see this, consider how the model might be applied to the experiment of Audley and Wallis that gave the crossover effect. At the perceptual stage the two bright lights would be coded as BRIGHT and BRIGHT+, and the two dim ones, as DIM and DIM+. At the linguistic stage the instruction "choose brighter" would be coded as BRIGHT+, and "choose dimmer" as DIM+. For the two bright lights (BRIGHT and BRIGHT+), then, it is a simple matter for the linguistic stage to match the instruction "choose brighter" (BRIGHT+) to the brighter of the two lights (BRIGHT+) and choose it as the correct response. It is not so simple, however, to match the instruction "choose dimmer" (DIM+) to one of the two lights BRIGHT and BRIGHT+, since neither of their codes is congruent with DIM+. Instead, the linguistic stage must translate the perceptual codes to DIM+ and DIM, respectively, and then make the match in order to select the correct response. Such a recoding is assumed to take time. Similar consequences follow from the application of the "choose dimmer" and "choose brighter" instructions to the pair of dim lights. Thus, the congruity effect arises at the linguistic stage from the additional time required to alter the perceptual code when it does not match the instructional code.

As just illustrated, the semantic congruity model can account for the crossover effect. It can also account for the funnel effect if we accept either of two assumptions. First, as suggested by both Audley and Wallis (1964) and Marks (1972), the response "brighter" could be enough faster than "dimmer," whatever the two lights judged, that the brighter judgment would always be faster than the dimmer judgment. So instead of the brighter and dimmer plots crossing over, they would form a funnel. Viewed this way, the funnel effect is simply a crossover effect with one of the two responses speeded up overall. This explanation of the funnel effect, by our model, would have to arise from operations subsequent to the linguistic stage—in response selection or execution, according to Smith's (1968) system. One weakness of this account, however, is that it does not say how different background illuminations lead to different response speeds, with one background giving the crossover effect and another, the funnel effect.

The second possible account puts the source of the funnel effect at the perceptual stage and remedies this weakness in the first account. Recall that the crossover effect arose in Audley and Wallis' experiment when the background was a middle gray, and the funnel effect arose when the background was virtually black. On a gray background, the dim pair appears dim compared to the background, and the bright pair appears bright compared to the background. On the black background, however, both pairs appear bright compared to the background: The dim pair appears only slightly bright, whereas the bright pair appears very bright. Given this rough description, a natural explanation for the funnel effect suggests itself.

The explanation goes like this. Assume that with a middle gray background the bright pair has a probability of .60 of being coded BRIGHT/BRIGHT+, and .40 of being coded DIM/DIM+; assume that the corresponding two values for the dim pair are .40 and .60, respectively. This pattern of probabilities reflects the assumptiton that the two pairs of lights are usually coded relative to the background, with the bright pair as BRIGHT/BRIGHT+ and the dim pair as DIM/DIM+. Given such a pattern, the operations of the linguistic stage would lead directly to a crossover effect. Now assume that the black background simply increases the probability of the BRIGHT/BRIGHT+ codes for both pairs, concomitantly lowering the probability of the DIM/DIM+ codes. For example, the probabilities of the BRIGHT/BRIGHT+ and DIM/DIM+ codes, respectively, might be .80 and .20 for the bright pair, and .60 and .40 for the dim pair. This pattern of probabilities is in agreement with the assumption that the two pairs of lights would both normally be coded as BRIGHT/BRIGHT+ relative to the black background, but differentially so for the bright and dim pairs. With such a pattern of probabilities, the operations of the linguistic stage would make all brighter choices faster than all dimmer choices, yet produce a congruity effect. The result would be the wanted funnel effect. The funnel effects in other experiments could result from similar coding preferences, though not necessarily preferences so purely under the control of stimulus factors.

Our aim in Experiment 1 was to show that the perceptual and linguistic stages of the present model are sequential and separate in function. We took two different tacks. First, we varied the discriminability of the two members of each pair of stimuli. When the two are very discriminable, the perceptual stage should be able to code the two very quickly; when they are not very discriminable, this stage should take much longer. Such changes in the discriminability, however, should only affect the speed of the perceptual stage, thereby advancing or delaying the start of the linguistic stage; they should not affect the linguistic stage directly. Thus, the overall reaction time should change with changes in discriminability, but the size of the congruity effect, which arises at the linguistic stage, should not. Second, we varied the instructions to the subject. One set of instructions was designed to produce a congruity effect at the linguistic stage, whereas another set of instructions was designed not to produce such an effect. The separation of the perceptual and linguistic stages would therefore be shown if the change in instructions proved to be additive to the changes in discriminability. Experiment 2 was designed to clear up a possible confounding in Experiment 1.

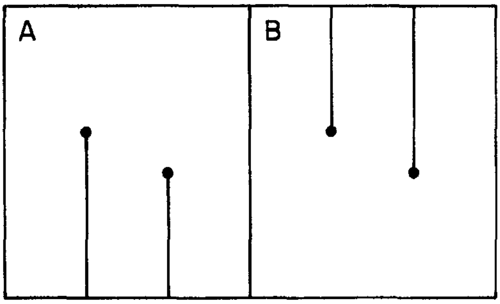

FIGURE 1. Examples of the stimuli used in Experiment 1: A, balloons; B, yo-yos.

We made use of some rather unusual stimuli: pairs of balloons (as shown in Figure 1A) and pairs of yo-yos (as shown in Figure 1B). We asked subjects either to choose the higher or lower of the pair, or to choose the longer or shorter of the strings. We chose these unusual objects because of how we thought subjects would code them for height and length. When asked to judge relative height, we assumed, the subjects would see the balloons as "going up" from the bottom and would ordinarily code them as HIGH and HIGH+. In contrast, we assumed, they would see the yo-yos as "going down" from the top and would ordinarily code them as LOW and LOW+. When shown the balloons, therefore, the subjects should be able to choose the higher faster than the lower, whereas when shown the yo-yos, they should find themselves doing just the reverse. This, of course, is the congruity effect, and it should appear as a proper crossover effect. But when asked to judge the relative lengths of the strings, we assumed, subjects would code the balloons and yo-yos in the same way. This is so because the length of a string does not depend on whether it "points" up or down. For both balloons and yo-yos subjects should code the strings as LONG and LONG+. Hence, the longer and shorter choices, unlike the higher and lower choices, should lead to no congruity effects for the balloons and yo-yos.

By this argument, the perceptual stage should generate identical length codes but different height codes for corresponding balloons and yo-yos, and these codes determine whether or not there is a congruity effect arising from the linguistic stage. To determine whether the behavior of these two stages was separate, we varied the discriminability of the two members of each pair. We accomplished this by varying the lengths of the strings attached to the balloons and yo-yos. When the two strings are nearly equal in length, the pair should not be very discriminable, and the perceptual stage should take considerable time to construct the code. When the two strings are quite unequal in length, the pair should become discriminable and should be coded more quickly. These variations at the perceptual stage should be additive to the congruity effects arising from the linguistic stage.

EXPERIMENT 1

Method

On each of a series of 136 trials, subjects were shown a question, such as "Which balloon is higher?", followed by a pair of appropriate stimuli, such as the pair of balloons in Figure 1A, and were required to press a button in answer to the question as quickly as possible while being timed.

The pairs to be judged consisted of 12 pairs of balloons and 12 pairs of yo-yos, as illustrated in Figure 1. The 12 pairs of balloons were constructed by pairing a string length of 2, 3, 4, or 5 cm for the left balloon with a string length of 2, 3, 4, or 5 cm for the right balloon, with the restriction that the two lengths could not be equal. The 12 pairs of yo-yos were constructed in the same way. The balloons and yo-yos themselves consisted of .2 cm diam. filled circles situated either 1.5 or .5 cm below, or .5 or 1.5 cm above the midline of a 13 × 7 cm display area. The only difference between the 12 balloons and the 12 yo-yos, therefore, was in whether the strings were drawn from the circles to the top edge of the 7-cm-high display, or from the circles to the bottom edge. The 2 balloons or 2 yo-yos of each display were separated horizontally by 2 cm. All were drawn in India ink.

The 12 pairs of balloons plus the 12 pairs of yo-yos were used in the construction of two sets of 48 displays, one set for height judgments and the other set for length judgments. The first set was formed by pairing each of the 24 pairs with each of the two questions "Which balloon [or yo-yo] is higher?" and "Which balloon [or yo-yo] is lower?". The second set was formed similarly by pairing each of the 24 pairs with each of the two questions "Which string is longer?" and "Which string is shorter?". One half of the subjects got the following sequence of 136 trials: (a) 20 practice trials chosen randomly from the 48 height displays; (b) the 48 height displays in an individually randomized order; (c) 20 practice trials chosen randomly from the 48 length displays; and (d) the 48 length displays in an individually randomized order. The other half of the subjects got the trials in the sequence *cdab*. When subjects erred on nonpractice trials, the erred-on display was placed farther down in the deck within the same block, and the trial was repeated.

The subject initiated each trial by pressing a "ready" button. The question appeared 1 sec later and remained on for 2 sec. Immediately after the question disappeared, the pair of balloons or yo-yos appeared and remained until the subject pressed the left or right button in a hand-held response panel to indicate the left- or right-hand member of the stimulus pair. The subject's RT was measured from the appearance of the balloons or yo-yos to his or her response. The subject viewed these displays in a modified Iconix tachistoscope. The question, typed in elite type, appeared in an upper 13 × 7 cm field, and the pair of balloons or yo-yos appeared in an identical 13 × 7 cm field 1 cm directly below the first. Between trials the subject viewed a blank 13 × 15 cm field that exactly covered the two smaller display fields. The subject viewed the displays at a distance of 51 cm.

The subjects were 12 Stanford University undergraduates who participated in order to fulfill a course requirement in introductory psychology. They were instructed to respond as quickly as possible, but to do so with as few errors as possible. The total session lasted about 50 min, with a short break in the middle of the experiment.

Results

Figure 2 shows the mean RTs for correct choices of the higher or lower balloon or yo-yo and of the longer or shorter string attached to the balloon or yo-yo. As this figure makes clear, there was a crossover effect for the higher/lower choices, but not for the longer/shorter choices. The crossover effect was highly reliable for the higher/lower choices alone, $F\ (1, 11) = 17.8$, $p < .005$; the slight funnel effect for the longer/shorter choices was not reliable, $F\ (1, 11) = 2.40$. The size of these two congruity effects, defined as the mean deviation of the two lines from parallel, was 40 msec with a standard error of 6.7 msec for the higher/lower choices, but only 12 msec with a standard error of 4.2 msec for the longer/shorter choices. Accordingly, the congruity effect was reliably larger for the higher/lower choices than for the longer/shorter choices, $F\ (1, 11) = 15.7$, $p < .005$.

Figure 2 shows several additional effects. First, the balloons were processed reliably faster than the yo-yos, by about 60 msec overall, $F\ (1, 11) = 33.7$, $p < .001$. Second, the advantage of balloons over yo-yos was reliably larger for the higher/lower choices than for the longer/shorter choices, $F\ (1, 11) = 10.0$, $p < .01$. And third, the longer/shorter choices were reliably faster than the higher/lower choices, F (1, 11) = 16.2, $p < .005$.

Figure 3 shows how both the higher/lower and longer/shorter choices got faster as the discriminability of the two balloons or yo-yos increased. Discriminability is expressed on the ordinate in Figure 3 as centimeters of vertical separation between the balloons or yo-yos. (Note that three conditions have been averaged to give each of the two points at 1 cm of discriminability, two to give each of the two points at 2 cm of discriminability, and one to give each of the two points at 3 cm of discriminability.) This decrease in RT with discriminability was highly reliable, F (5, 55) = 13.6, $p < .001$, and the decrease was statistically parallel for the higher/lower and longer/shorter choices ($F < 1$) which differed by about 50 msec at each level of discriminability.

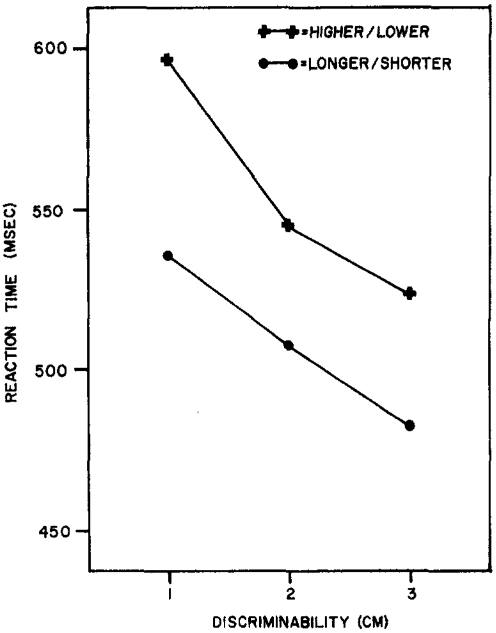

Figure 3. Reaction times for the higher/lower and the longer/shorter choices, plotted as a function of discriminability defined as the difference in cm between the lengths of the strings (Experiment 1).

What is crucial for the semantic congruity model is that these discriminability effects be additive to the congruity effect shown earlier. That is, discriminability should not interact statistically with the congruity effect, and indeed it did not. For example, the interaction of discriminability with the pattern shown in Figure 2 (crossover effect for the higher/lower choices and none for the longer/shorter choices) was statistically unreliable ($F < 1$). This lack of significance does not seem to have arisen simply from noisy RTs, for the standard error of this interaction was only 6.9 msec. This interaction computed separately for the higher/lower and longer/shorter choices was also unreliable in both cases, and the standard errors were 10.6 msec for higher/lower and 7.9 msec for longer/shorter.

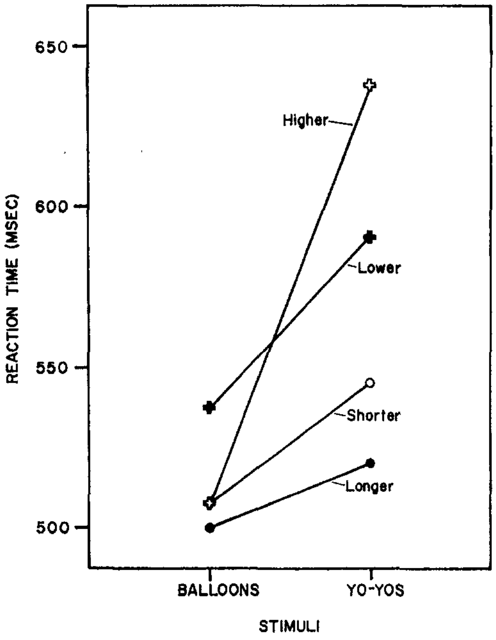

Figure 2. Reaction times to choose either the higher or lower balloon or yo-yo, or the longer or shorter balloon or yo-yo string (Experiment 1).

To illustrate this additivity more graphically, Figure 4 shows the crossover effect of the higher/lower choices for the three levels of discriminability separately. The additivity of these two effects in this plot is striking. The size of the crossover effect is 44, 31, and 42 msec for discriminability levels of 1, 2, and 3 cm, respectively. In arithmetic terms, the additivity of the effects in

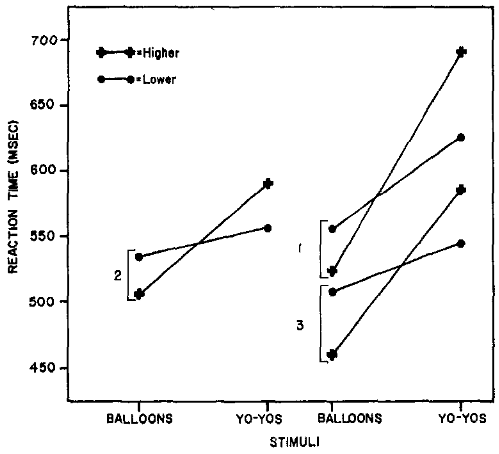

FIGURE 4. Crossover effects for the higher/lower choice at the three levels of stimulus discriminability (Experiment 1).

Figure 4 can be stated as follows: One must move each point an average of 31 msec to remove the effect of discriminability, 20 msec to remove the crossover effect, and only 6.4 msec to remove the remainder (the nonadditivity between discriminability and the crossover effect). Since the standard error of this interaction is 10.6 msec, the nonadditivity of 6.4 msec is well within the range of error. Furthermore, the nonadditivity is not monotonic with discriminability and hence does not hint at a systematic violation of the model. Finally, the effect of discriminability is roughly parallel for balloons and yo-yos: As discriminability increases, the absolute higher/lower differences are 31, 25, and 45 msec for balloons and 58, 38, and 40 msec for yo-yos. Thus, averaging the two types of objects does not obscure violations of the model, and the data, overall, are consistent with our additive model.

The mean error rate was 3.0%, ranging from 0% to 6.8% for individual subjects. We did not record the RTs for errors. Although there were too few errors for detailed analysis, they generally covaried with mean RTs over conditions, with a product-moment correlation of .35 for the longer/shorter choices and .28 for the higher/lower choices. In addition, it made no difference to RTs whether the subjects made the higher/lower choices before or after they made the longer/shorter choices. None of the relevant interactions with order were reliable.

EXPERIMENT 2

In Experiment 1, the higher/lower choices showed a semantic congruity effect, but the longer/shorter choices did not. Unfortunately, this critical result could have arisen from a simple artifact of the configurations of balloons and yo-yos. Note that subjects were faster in choosing the higher balloon and the lower yo-yo; that is, they were faster in choosing the balloon with the longer string and in choosing the yo-yo with the longer string. Thus, string length was confounded with speed of choosing the correct object. This raises the following possibility. The subjects could actually have been making their higher/lower choices by comparing string lengths instead of relative heights; and, of course, they would have been faster at choosing the longer string, just as they were in the other half of the experiment. If the congruity effect of the higher/lower choices could be explained away in this manner, then we would have literally nothing in our data to explain with the two-stage model.

We designed Experiment 2 to rule out this unpleasant alternative. The idea, essentially, was to rerun the higher/lower choices of Experiment 1, but with all of the strings on the balloons and yo-yos of equal length. Hence the subjects could not possibly use string length in making their higher/lower choices. If the semantic congruity effect emerged in Experiment 2, and with equal magnitude, then we could rule out this alternative explanation for the higher/lower choices of Experiment 1.

Method

On each of 168 trials subjects were shown one of four questions ("Which balloon/yo-yo is higher/lower?") followed by a pair of balloons or yo-yos, whichever was appropriate, and were required to press a button in answer to the question as quickly as possible while being timed.

The set of 24 displays used in Experiment 2 was identical to the set of 48 higher/lower displays used in Experiment 1 with two exceptions. First, the strings attached to all balloons and yo-yos in Experiment 2 were only 1 cm long. Second, Experiment 2 used only those 24 displays of Experiment 1 that had a horizontal discriminability of 1 cm; the remaining 24 displays of Experiment 1 with discriminabilities of 2 and 3 cm were omitted. Each subject was presented this set of 24 displays seven times, each time in a different individually randomized order. The first block of 24 trials was considered practice. In all other respects, the procedure, equipment, and instructions were the same as in Experiment 1, and the 14 new subjects were from the same source as in Experiment 1.

Results

Table 1 shows the mean correct RTs for choosing either the higher or lower balloon or yo-yo. These RTs show a clear crossover effect, just as in Experiment 1, and it is highly reliable, F (1, 13) = 32.6, $p <$ .001. The size of the crossover effect was 35 msec, with a standard error of 4.4 msec. Furthermore, there was a full crossover effect at all three pairs of positions of the two balloons or yo-yos, and the size of the effect did not vary reliably with position ($F < 1$), where the standard error of this interaction was 6.6 msec.

Apparently, therefore, the alternative "string length" explanation for Experiment 1 cannot be correct. First, there was a full crossover effect in Experiment 2, where all the string lengths were identical. There was no way string length could have produced this crossover effect. And second, the size of the crossover effect in Experiment 2 (35 msec) was very nearly the same as in Experiment 1 (40 msec), the difference between them falling far short of statistical significance ($F < 1$). So the hypothesized "string length" effect does not appear to contribute at all to the size of the crossover effect.

TABLE 1

Mean Reaction Times (in msec) for Choosing the Higher or Lower Balloon or Yo-Yo in Experiment 2

Instruction	Pictured objects	
	Balloons	Yo-yos
Which is higher?	535 (.2)	608 (1.4)
Which is lower?	562 (1.6)	565 (2.6)
M	548	586

Note. The percentage of errors is given in parentheses.

There were two other results of interest in Experiment 2. First, subjects were faster at judging balloons than yo-yos, F (1, 13) = 19.1, $p <$ .005, just as they were in Experiment 1. The size of this advantage was reliably smaller in Experiment 2 (38 msec) than in Experiment 1 (91 msec), F (1, 24) = 26.31, $p <$ .001. The smaller advantage in Experiment 2 could have been caused by perceptual strategies associated with the shorter strings in Experiment 2. This result, in any event, is in accord with the semantic congruity model, since the congruity effect, the result of a later stage of processing, was unaffected by the change brought about by this perceptual strategy.

In Experiment 2 (unlike Experiment 1), we recorded the side of the correct response (left or right) and found the right side 21 msec faster than the left, F (1, 13) = 13.9, $p <$ .005. Since our subjects always used their right hand to choose the right side, and their left hand the left side, we cannot know whether this 21-msec difference resulted from a dexterity advantage in the right hand or from a preferred order in visual scanning. However, the right side was 53 msec faster than the left side for the "choose higher" instruction and 12 msec slower than the left side for the "choose lower" instruction. This interaction was reliable, F (1, 13) = 7.2, $p <$ .025. We can only speculate on the source of this phenomenon. Perhaps the positive instruction "choose higher" facilitates the response compatible with a positive code, the right hand, and the negative instruction "choose lower" facilitates the opposite response, the left hand. Whatever its source, this phenomenon interacted neither with the congruity effect ($F < 1$) nor with any other effect in the experiment.

The error rate in Experiment 2 was only 1.4%, with individual subjects varying from 0% to 5.6%. Error RTs were not recorded. Again there were too few errors for a detailed analysis (29 in all), but their distribution is shown in Table 1.

DISCUSSION

The results of Experiments 1 and 2 are consistent with the two basic assumptions of the semantic congruity model. First, stimulus discriminability and linguistic congruity led to additive effects in Experiment 1. This finding is consistent with the assumption that the perceptual stage occurs entirely prior to the linguistic stage. Second, the higher/lower choices produced a congruity effect, and of about the same size in Experiments 1 and 2, whereas the longer/shorter choices did not. This finding supports the notion that the congruity effect arises at the linguistic stage. In addition, the results appear to rule out several previous explanations of the congruity effect.

The Semantic Congruity Model

In the semantic congruity model, the perceptual stage comes first, generating perceptual codes based on the two stimuli being compared. By assumption, it generates these codes more quickly the more discriminable the two stimuli are. Experiment 1 supported this assumption. As shown in Figure 3, the subjects got faster in their comparative judgments the greater the vertical separation of the two balloons or yo-yos. This result, however, is hardly new or surprising (see Smith, 1968; Welford, 1960).

Under this model, the perceptual stage may also account for several other findings in these experiments. Under the higher/lower instructions, subjects were much faster (92 msec) on balloons than on yo-yos; under the longer/shorter instructions, they were also faster on balloons, but by only a slight amount (28 msec). The large 92-msec difference under the higher/lower instruction may have occurred because subjects were able to generate perceptual codes faster for balloons than for yo-yos. This would follow from our assumption that subjects coded the balloons as the unmarked, and easier, HIGH/HIGH+, and the yo-yos as the marked, and harder, LOW/LOW+.[1] On the other hand, under the longer/shorter instructions, subjects were assumed to code both the balloons and the yo-yos as LONG/LONG+, and so the difference between balloons and yo-yos should be less, just as was found.

But why were subjects still faster on balloons than on yo-yos when they coded both of them as LONG/LONG+? To be consistent with the model, this 28-msec difference must also arise at the perceptual stage, for once subjects have coded two objects, the processing times at later stages should not depend on whether the objects coded were balloons or yo-yos. Although the model is not detailed enough to say just why balloons were coded faster than yo-yos, it may be that subjects have visual scanning strategies or expectations that favor the balloons. If so, the 92-msec advantage of balloons over yo-yos could be argued to arise from two separate effects, both perceptual: (a) Yo-yos take 28 msec longer than balloons because they take longer to scan or because they are less expected; and (b) the yo-yos in the higher/lower condition take an additional 64 msec longer because their LOW/LOW+ codes take longer to form than the simpler HIGH/HIGH+ codes for the balloons. We note, however, that the latter 64-msec difference could instead be attributed to the linguistic stage, which could find it easier to compare the instructional code with HIGH/HIGH+ than with LOW/LOW+. At present, we favor the earlier interpretation, but only because we can think of no compelling reason why comparisons with HIGH/HIGH+ should be any faster than those with LOW/LOW+.

Operations at the perceptual stage may account also for why subjects were faster in the longer/shorter conditions than in the higher/lower conditions. In the longer/shorter conditions, subjects had but one pair

---

[1] On the basis of several linguistic criteria, the adjective pairs *long–short* and *high–low*, like many other such pairs, consist of a positive, or "unmarked," member (*long* and *high*) and a negative, or "marked," member (*short* and *low*) (Clark, 1969). Unmarked adjectives have been shown to have simpler structure, and they take less time to code and compare than their marked antonyms (see Clark, 1969; Clark, Carpenter, & Just, 1973).

of codes for all objects, namely, LONG/LONG+. In the higher/lower conditions, however, they had to choose between HIGH/HIGH+ and LOW/LOW+, depending on the objects. This choice may itself have increased the coding times in the higher/lower conditions. Nevertheless, this finding could also be attributed to the linguistic stage, for the LONG/LONG+ codes may somehow be easier to deal with in comparisons than the HIGH/HIGH+ and LOW/LOW+ codes.

The linguistic stage comes after the perceptual stage. It compares the perceptual code with the previously stored instructional code and selects the correct response. The duration of this stage is the sum of the times taken by the obligatory comparison operations plus the times taken by the additional transformations that are sometimes needed to bring the perceptual and linguistic codes into congruence. The congruity effect then follows naturally. In brief, when the direction of the judgment (higher and lower) is the same as the stimuli as coded (balloons or yo-yos), the codes will already be congruent and so the RTs will be shorter. This is just what we found. No congruity effect, however, should arise for the longer/shorter choices because the codes are assumed to be the same for both the balloons and the yo-yos. This too is just what we found. Finally, the fact that the longer choice was faster than the shorter choice (by 17 msec) is consistent with the assumption that the strings are usually coded as LONG/LONG+. These codes match the instructional codes for the "choose longer" instruction, but not for the "choose shorter" instruction. According to the model, then, the longer/shorter difference is a type of semantic congruity effect.

In the semantic congruity 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, of course, because the perceptual code would always have matched the instructional code, and the linguistic stage would never have required an extra, time-consuming step. According to the present account, however, subjects did not set themselves in this fashion, for if they had, they would not have shown a congruity effect. Instead, they were forced to code the balloons as "going up" and the yo-yos as "going down," apparently because setting themselves to do otherwise was impossible. This is consistent with a review of the previous work on comparative judgments by Clark et al. (1973).

Without changing most of its basic assumptions, however, we could replace the semantic congruity model with a related "perceptual congruity model" that would incorporate this optimal strategy for coding the balloons and yo-yos. In this model, the principle of congruence would operate not at the linguistic stage, but at the prior perceptual stage. When asked to choose the higher object, the subject would code the two objects—whether balloons or yo-yos—as HIGH/HIGH+, and when asked to choose the lower object, he would code them as LOW/LOW+. The assumption, however, would be that he would independently try to code the balloons as HIGH/HIGH+, and the yo-yos as LOW/LOW+. He would therefore be helped in forming the HIGH/HIGH+ codes for balloons, and the LOW/LOW+ codes for yo-yos, and he would be interfered with, and slowed down, in forming the alternate codes. This view of perceptual coding still attributes an inflexibility to the perceptual coding of balloons and yo-yos, but the inflexibility only results in the slower formation of codes that conflict with the "natural" codes. The perceptual congruity model, however, would not predict a lack of interaction between the discriminability and congruity effects unless it had separate, sequential substages for identifying the objects and coding their comparative heights. For now, we favor the semantic congruity model because it clearly separates the discriminability and congruity effects by placing them at separate stages. Future work, however, may require a more complex view of the perceptual stage and revive this issue.

Alternative Models

The present findings may also be used to evaluate several earlier explanations of the crossover effect and related semantic congruity effects. There are three principal models to be examined. Audley and Wallis (1964) based an explanation of the congruity effect on Audley's (1960) stochastic model of choice behavior. In a subsequent paper, Wallis and Audley (1964) presented a second model based on response competition. And finally, Marks (1972) proposed a model based on the discriminal dispersions of the stimuli on the underlying continuum of judgment.

The stochastic choice model. According to Audley's choice model, the subject is implicitly producing each possible response in a choice situation at some rate. When he finds a sufficiently long uninterrupted run of one of the implicit responses, he executes the corresponding overt response. For example, when a subject must decide whether a single fairly bright light is dim (D) or bright (B), he produces a series of implicit responses such as BDBBDBDBBBDBBBB. When the light is made brighter, the Bs come faster, the probability of a criterial run of Bs (say, five Bs) becomes greater, and he is quicker to respond "bright." Working from this model, Audley and Wallis (1964) proposed the following account for the crossover effect. When a subject is asked to choose the brighter of two bright lights, he must wait for a criterial run of Bs from one of the two lights; when he is asked to choose the dimmer of these same lights, he must wait for a criterial run of Ds. Since the Bs occur more frequently than the Ds for these two bright lights, he reaches criterion much more quickly for the Bs than for the Ds, and so the brighter choice is made faster than the dimmer one. For the two dim lights, the situation is just the reverse, and the result is the crossover effect. By similar reasoning, the stochastic choice model (Audley, 1960, p. 9) also accounts for the decrease in RT with increases in discriminability of the two lights (see Figure 3).

The stochastic choice model, however, also predicts that the crossover effect should vary systematically with discriminability. Note first that the brighter of two bright lights produces more Bs than the dimmer one produces Ds; therefore, the brighter choice should be faster than the dimmer choice. But making the brighter light brighter while leaving the dimmer one constant should lead to even more Bs and an even faster brighter choice, but to no more Ds and so no faster dimmer choice. The difference between the brighter and dimmer choices should therefore increase. By similar reasoning, making the dimmer of two bright lights dimmer while leaving the other constant should *decrease* the difference between the brighter and dimmer choices.

These predictions can be checked against the present results. First, for those pairs of balloons containing the lowest balloon, and for those pairs of yo-yos containing the highest yo-yo, the absolute higher/lower difference should increase with the three steps of increasing discriminability. But the appropriate numbers are 52, 118, and 85 msec, a nonmonotonic increase with only 33 msec between the extremes. Second, for those pairs of balloons containing the highest balloon, and for those pairs of yo-yos containing the lowest yo-yo, the absolute higher/lower difference should decrease with the three steps of increasing discriminability. Here, the appropriate numbers are 101, 15, and 85 msec, again a nonmonotonic trend but with only 16 msec between the extremes. Finally, we can examine the trend for those pairs containing the next-to-lowest balloon and the next-to-highest yo-yo; with the two steps of increasing discriminability, they should show increasing numbers. But the actual numbers are 112 and 15 msec, a trend in the wrong direction.

Although these tests cannot be taken as strong disconfirmation of the stochastic choice model, they clearly provide little support. The three trends were either strongly nonmonotonic or in the wrong direction, and the differences between the extremes in the first two tests (33 and 16 msec, respectively) were relatively small compared to the 92-msec variation in discriminability in these extreme cases and compared to the 98-

msec trend in the wrong direction for the third test.

Response competition model. Wallis and Audley (1964), however, suggested another possible model for the crossover effect, one they called the response competition model. According to this model, the subject produces the same implicit responses as in the previous model. But here he uses the most prevalent response (B, for example, if the two lights are bright) as an implicit choice, and he gives this response if it matches the instruction. He must change the response to its alternative if the instruction requires the opposite response. So when asked for the brighter of two bright lights, he can respond directly and quickly; but when asked for the dimmer of the same two lights, he must translate and respond more slowly. This reasoning would lead to the congruity effect.

From Audley and Wallis' one-sentence description, the response competition model does not seem very different from the present semantic congruity model. Indeed, it seems somewhat inconsistent with their previous assumption that the implicit responses—the Bs and Ds associated with the stimuli—serve as the sole basis for the overt responses, since in the response competition model they do not. In our own model we have simply tried to be more explicit about the two stages implicit in their model. We have given evidence the two stages are sequential, and we have provided explicit accounts for the linguistic codes and comparison operations. In addition, our model provides a natural explanation for the effects of linguistic markedness noted in the present results, and the principle of congruence provides our model with some generality since it links our model with other models of linguistic processing.

The discriminal dispersion model. Marks (1972) has recently proposed quite a different explanation for the congruity effect. According to his discriminal dispersion model, the subject codes each stimulus using one of the two polar opposite labels that are appropriate to the dimension on which the stimulus lies. For example, a bright light could be coded on either a bright continuum or a dim continuum. But the critical assumption of this model, as applied to these two continua, is that as a light gets brighter its variance on the bright continuum gets smaller, while its variance on the dim continuum gets larger. Then to predict RTs for the questions "Which is brighter?" and "Which is dimmer?" for a given pair of lights, Marks assumes that the RT for the former judgment, RT_B, is given as in Equation 1, and the RT for the latter, RT_D, is given as in Equation 2:

$$RT_B = C_1\sigma_B^2/d, \qquad (1)$$

$$RT_D = C_2\sigma_D^2/d. \qquad (2)$$

Here, σ_B^2 is the average variance of the two lights on the bright continuum, and σ_D^2 is their average variance on the dim continuum; C_1 and C_2 are proportionality constants; and d is the absolute difference between the two lights. For a pair of very bright lights, σ_B^2 will be small, and σ_D^2 will be large, and so with C_1 and C_2 approximately equal, RT_B will be shorter than RT_D. For a pair of very dim lights, σ_B^2 will be *larger* than σ_D^2, and so it is the RT_D that will be shorter. This pattern of RTs, of course, is the familiar crossover effect. With other values for C_1 and C_2 the RTs will result in the funnel effect, or in some other congruity effect.

Although this model is successful in predicting the congruity effects, it also predicts that the size of these congruity effects increases with discriminability. Consider two moderately bright lights. If C_1 and C_2 are about equal, RT_D (the RT for the dimmer choice) should be considerably larger than RT_B (the RT for the brighter choice), and so M, the value of $RT_D - RT_B$, should be positive:

$$M = (C_2\sigma_D^2 - C_1\sigma_B^2)/d. \qquad (3)$$

But this equation shows that M will decrease as d increases, so long as the average variances σ_D^2 and σ_B^2 remain approximately constant. The value of M for very dim lights, likewise, will be negative and should decrease in absolute size as d increases. Consequently, the size of the congruity effect should decrease as the discriminability of the two lights increases. In the present experiment, where *d* is the same for balloons and yo-yos, the congruity effect should vary about 2 msec for every 1 msec change in discriminability, assuming the σ^2's remain about the same. But, as shown in Figure 4, the size of the congruity effect remained substantially the same over a wide variation in discriminability.

In his own discussion of the congruity effect, Marks pointed out:

> While the discriminal dispersion theory and Clark's (1969) linguistic theory [roughly equivalent to the present semantic congruity theory] are not inconsistent, neither are they identical. For example, Clark's theory implies that relative judgment is a serial multistage process, while the discriminal dispersion theory gives a single-stage model. Further data are needed to determine which theory is the better. (p. 160)

The present results appear to provide the critical data, and they support the semantic congruity model, a multistage theory, over the discriminal dispersion model, a single-stage model.

We have demonstrated that a two-stage model for the semantic congruity effect is able to account for the existing data quite nicely. There are, however, a number of details missing from this model. At present we know very little about the perceptual stage other than that discriminability has its effects there and so may certain other perceptual factors. Exactly what other factors have an effect there, and why, remains to be seen. We also know very little about the workings of the linguistic stage. Apparently, markedness has its effects on this stage, and the comparison process operates in accordance with the principle of congruence, but as yet we do not know how various other instructions may be coded there and compared. Experiments like this do show, however, that what psychophysicists always thought were rather simple judgments, depending only on the properties of the pair of stimuli being compared, are not simple at all. The instructions, simple as they are, have their effects on this task too, and the "properties" of the stimuli are really perceptual interpretations.

REFERENCES

Audley, R. J. A stochastic model for individual choice behavior. *Psychological Review,* 1960, *67,* 1–15.

Audley, R. J., & Wallis, C. P. Response instructions and the speed of relative judgments. I. Some experiments on brightness discrimination. *British Journal of Psychology,* 1964, *55,* 59–73.

Clark, H. H. Linguistic processes in deductive reasoning. *Psychological Review,* 1969, *76,* 387–404.

Clark, H. H., Carpenter, P. A., & Just, M. A. On the meeting of semantics and perception. In W. G. Chase (Ed.), *Visual information processing.* New York: Academic Press, 1973.

Ellis, S. H. Interaction of encoding and retrieval in relative age judgments: An extension of the "crossover" effect. *Journal of Experimental Psychology,* 1972, *94,* 291–294.

Friend, K. E. Perceptual encoding in comparative judgments of race. *Memory and Cognition,* 1973, *1,* 80–84.

Marks, D. F. Relative judgment: A phenomenon and a theory. *Perception & Psychophysics,* 1972, *11,* 156–160.

Shipley, W. C., Coffin, J. I., & Hadsell, K. C. Reaction time in judgment of color preference. *Journal of Experimental Psychology,* 1945, *35,* 206–215.

Shipley, W. C., Norris, E. D., & Roberts, M. L. The effect of changed polarity of set on decision time of affective judgments. *Journal of Experimental Psychology,* 1946, *36,* 237–243.

Smith, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin,* 1968, *69,* 77–110.

Wallis, C. P., & Audley, R. J. Response instructions and the speed of relative judgments. II. Pitch discrimination. *British Journal of Psychology,* 1964, *55,* 133–142.

Welford, A. T. The measurement of sensory-motor performance: Survey and reappraisal of *twelve years' progress. Ergonomics,* 1960, *3,* 189–230.

(Received June 10, 1974)