Semantic congruity effects in perceptual comparisons

WILLIAM M. PETRUSIC and JOSEPH V. BARANSKI
Carleton University, Ottawa, Ontario, Canada

Semantic congruity effects occur when, as in Experiment 1, for example, the time to select the shorter of two relatively short lines is faster than the time to select the longer; conversely, selection of the longer of two relatively long lines is faster than selection of the shorter. Semantic congruity effects are also demonstrated in experiments requiring comparisons of the heaviness of weights (Experiment 2) and horizontal extent (Experiment 3). In Experiment 1, the magnitude of the semantic congruity effect was larger under conditions emphasizing accuracy rather than speed and when the comparison was difficult. In fact, when comparisons were errorless, the effect was minimal (20 msec), thereby replicating previous failures to obtain the effect with supraliminal perceptual comparisons (Banks, Mermelstein, & Yu, 1982; Marschark & Paivio, 1981; Petrusic & Baranski, 1988a). In confirmation of Henmon's (1911) introspective analyses of psychophysical comparisons, Experiments 2 and 3 extend the range of the semantic congruity effect to include judgments of confidence. However, in each of the three experiments, semantic congruity effects were not evident with the response-accuracy measure. Finally, using highly confusable stimuli, in Experiment 3 the magnitude of the semantic congruity effect was shown to be larger for error than for correct response times. The implications of these findings for a *decisional locus* of the effect and for the semantic coding theory are discussed.

Henmon (1911) observed that when subjects identified the "longer" or the "shorter" of two relatively short lines

one curious constant error in judgments of the shorter line appeared in the results. All of the subjects . . . noted . . . that judgment could be given more easily, more quickly and with greater confidence when reaction was made to the shorter line. The feeling that the most accurate judgments would be secured with the shorter line was very marked. (Henmon, 1911, pp. 194-195)

Henmon's introspective observations met with remarkably little empirical support in the very extensive experimental data he obtained from his four observers (including himself). In the present article, however, experimental findings that confirm and extend Henmon's observations will be presented. We now know that if Henmon had repeated his experiment with long lines he would have obtained the opposite findings: selection of the longer

member of the pair would have been made faster and with more confidence.

In 1964, working with the brightness continuum, Audley and Wallis found that the selection of the brighter of two bright lights was faster than the selection of the darker. Conversely, selection of the darker of two dark lights was faster than the selection of the lighter. Audley and Wallis (1964), who brought the phenomenon to the attention of contemporary experimentalists, surmised that the effect might represent a general principle of comparative judgment, and they referred to this interaction between the direction of the comparison and the location of the alternatives on the continuum as the *crossover effect*.¹ Thus Wallis and Audley (1964) provided an immediate replication of the effect: using pure tones varying in pitch, they showed the effect to also occur with successive comparisons of stimuli varying on a metathetic continuum.

Since the Audley and Wallis (1964) and Wallis and Audley (1964) experiments, the same effect has been found with a variety of perceptual continua. Ellis (1972) and Friend (1973) found the effect for pictorial comparisons of relative ages and skin colors. Likewise, Banks and Root (1979) found congruity effects for judgments of loudness, and Holyoak and Patterson (1981) found the effect for the relative ordering of colored sticks in a visual array. It should be noted, however, that few of these experiments required that the comparisons involve magnitudes of stimuli varying on *strictly perceptual* continua. For example, it is not clear that comparisons of the ages of people in photographs involve exclusively perceptual information.

This research has been supported by grants from the Natural Science and Engineering Research Council of Canada to William M. Petrusic and graduate research support from the Department of Psychology, Carleton University, to Joseph V. Baranski. Portions of this work were presented by Petrusic at the 1983 meeting of the Psychonomic Society and at the 1984 and 1985 meetings of the Canadian Psychological Association. The data in Experiment 2 arise from reanalyses of data originally collected by Nancy Lochnan in Petrusic's laboratory as part of an MA thesis submitted to the Department of Psychology, Carleton University, 1977. We are grateful to Lester Krueger, Donald Jamieson, and Bill Jones for helpful comments. Requests for reprints should be addressed to William M. Petrusic, Department of Psychology, Carleton University, Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada.

In addition to being reliable, the effect has also proven to be robust, especially when judgments of the properties of remembered (or symbolically coded) alternatives are required. For example, Banks, Clark, and Lucy (1975) showed that comparisons of the height of appropriately labeled perceptual displays (strings coded either as yo-yos or as balloons) yielded the congruity effect. At the same time, Jamieson and Petrusic (1975c), extending Moyer's (1973) "internal psychophysics," demonstrated a fully graded crossover effect that increased regularly with extremity, by using comparisons of the remembered size of animals. As well, Holyoak and Walker (1976) provided additional demonstrations of the effect with stimuli varying on several symbolic continua (see Banks, 1977; Kosslyn, 1980; Moyer & Dumais, 1978 for reviews of congruity effects with symbolic comparisons).

In line with the latter findings, two theories proposed the exclusive use of semantic (associative priming) or symbolic (memorial) information as the basis for deriving the congruity effect. Depending on where in the comparison process the effect is thought to occur, they can be dichotomized as being primarily either predecisional or postdecisional in locus.

A predecisional locus is assumed in several versions of an *expectancy* hypothesis (Kosslyn, Murphy, Bemesderfer, & Feinstein, 1977; Marschark & Paivio, 1979; 1981; Moyer & Dumais, 1978; Shipley, Norris, & Roberts, 1946). According to this view, the direction of the comparison in some way prepares the subject for a range of stimuli that is congruent with that comparison. If they are, the response can be given immediately. If not, additional time is required to accommodate the range, thus giving rise to the congruity effect.

A somewhat more detailed model of the semantic congruity effect, proposing a postdecisional locus, is the *semantic coding theory* (Banks, 1977; Banks et al., 1975; Banks, Fujii, & Kayra-Stuart, 1976; Banks, Mermelstein, & Yu, 1982). According to this view, the effect arises as a consequence of discrete stimulus and instructional code matches or mismatches. For example, two small items are coded as "small" and "small+" and two large items are coded as "large" and "large+" (the + indicates the extreme along the dimension of the stimulus pair). When the instructional code matches the stimulus codes, the overt response can be immediately performed. However, when the instructional code and stimulus codes do not match (i.e., the stimuli are "small" and "small+" and the instruction says to choose the larger), a code translation is required. This additional time requirement, when stimuli are incongruent with the instruction, gives rise to the semantic congruity effect. A postdecisional locus of the effect is thus assumed, since code translation, which accounts for the congruity effect, occurs after the discrimination (code formation) is completed.

The semantic coding and the expectancy theories both contributed to reinstate an interest in the occurrence of the semantic congruity effect in purely perceptual comparisons. In fact, Marschark and Paivio (1981) did a series of experiments on the basis of their expectancy the-

ory, in order to question the existing semantic congruity effects with perceptual stimuli; generally, they predicted that these effects would not occur. They reported five experiments in which they failed to find the effect with "pure" perceptual comparisons (visual size judgments), and they concluded that these failures indicated that associative priming is impossible with strictly perceptual continua. They state,

The present experiments, together with the majority of the studies in the relevant literature, indicate that congruity effects are confined to tasks that involve symbolic, memorial information as the basis for comparative judgments. These findings suggest associative priming as the locus of congruity effects, a proposal consistent with the failure to obtain perceptual congruity effects. (Marschark & Paivio, 1981, p. 307)

Marschark and Paivio attributed the demonstrations of Friend (1973), Ellis (1972), and Banks et al. (1975) to the use of symbolic knowledge of facial and or physical characteristics, and they attributed the classic demonstrations of the effect by Audley and Wallis (1964) and Wallis and Audley (1964) to the use of a limited number of stimuli and, consequently, to "uninteresting consequences of perfect or near perfect memory or the operation of imaginal mediational processes" (Marschark & Paivio, 1981, pp. 294-295).

Since then, several experiments have provided for the clear rejection of the expectancy theory of the semantic congruity effect. Banks and Flora (1977), Banks, White, Sturgill, and Mermelstein (1983), and Holyoak and Mah (1981) presented evidence refuting the single clear prediction from the expectancy theory—that is, that the effect should not occur when the instructions specifying the form of the comparison follow stimulus presentation. Banks et al. (1983) also provided a direct test of the expectancy theory based on associative priming by intermixing lexical decision-making with comparison trials and comparing the magnitude of the lexical priming effect with that of the semantic congruity effect. They failed to find a systematic relationship between semantic congruity magnitude and the amount of facilitation on the lexical decision task. However, perhaps the most conclusive evidence against the expectancy theory was provided by Howard (1983). Before each trial, he provided subjects with the precise location of each stimulus pair on the continuum, thereby overriding any possible expectancy effects. A large congruity effect was obtained nevertheless.

Given these rejections of the expectancy theory, the problem remains as to why Marschark and Paivio (1981) failed to find semantic congruity effects with their perceptual stimuli. The failure to find the effect could be entirely methodological, or it could be reliable and in need of an alternative theoretical basis. Indeed, Banks et al. (1982), following Banks' (1977) semantic coding theory, state explicitly that

comparative judgments among perceptual stimuli will show a congruity effect only when they are encoded and processed in terms of an underlying semantic continuum. That is, the

congruity effect will be found only when the physical stimuli are interpreted symbolically. (p. 267)

In support of their position, Banks et al. (1982) failed to find statistically reliable evidence of semantic congruity effects in a circle-size comparison task. Moreover, semantic congruity effects were clearly obtained when these stimuli were compared in memory after subjects had learned to associate a nonsense syllable with each circle. More recently, Petrusic and Baranski (1988a) provided a replication (with circle size) and an extension of the Banks et al. (1982) findings, showing clear failures of the occurrence of the effect with, in this case, comparisons of vertical extent that spanned a very wide range of the extent continuum. Thus, contrary to the original reports of reliable and substantial effects with perceptual stimuli reported by Audley and Wallis (1964) and Wallis and Audley (1964), it is clear that the semantic congruity effect has also not been obtained in a wide variety of different experiments.

However, a review of the properties of the stimuli in the experiments in which semantic congruity effects with strictly perceptual comparisons were not obtained will show each experiment to have involved supraliminally different stimuli—that is, stimulus pairs with perfect or near perfect discriminability. On the other hand, in the Audley and Wallis (1964) and Wallis and Audley (1964) experiments, where the effect was obtained, the error rate was not specified. Perhaps, the effect is not obtained only when performance is errorless or nearly errorless.

The present experiments were conducted with three major purposes in mind. First, to test the proposition that semantic congruity effects can occur in perceptual comparisons, stimulus pairs with a sufficiently high error rate on each of three perceptual continua, not previously explored, were examined for congruity effects. In addition, the use of a relatively large number of pairwise *confusable* stimuli, sampled over a very wide range on each continuum, minimized possible problems of memory of a particular location and the correct alternative at that location. Second, since pairs with a sufficiently high error rate were included in the stimulus sets used, these experiments permitted extension of the investigation of the semantic congruity effect to determine the degree to which response accuracy (the probability of correct responses) is subject to the semantic congruity effect. Finally, with a view toward determining the basis for the effect and providing preliminary tests of the available theories, the roles of variables known to exert their influences primarily on the decision process were examined. In particular, the possible dependence of the magnitude of the semantic congruity effect on the relative demands for speed and for accuracy as well as on the difficulty of the decision was studied.

EXPERIMENT 1

Experiment 1 was an attempt to determine if semantic congruity effects occur in the classic psychophysical line-

length comparison task. We included stimulus pairs that were easy to discriminate (errorless) a priori, and pairs that required very difficult discriminations. Consequently, it was possible to determine if the magnitude of the effect was indeed dependent on the difficulty of the comparison. Also, as indicated above, we used a relatively large number of pairwise confusable stimuli, located throughout the attribute, in order to minimize the possibility that the coding of the location of the elements of the stimulus pair on the attribute could mediate the discrimination itself. The occurrence of semantic congruity effects with highly confusable stimuli would thereby weaken the argument that perceptual congruity effects occur only if mediated by memorial, imaginal, or symbolic interpretation (Banks et al., 1982; Marschark & Paivio, 1981)—although of course the possible occurrence of semantically mediated perceptual comparison processes could not be excluded in principle. Finally, we explored the possible effects of the relative demands for speed and accuracy on the magnitude of the semantic congruity effect.

Method

Subjects. Four graduate students in psychology served as volunteers for 12 1-h sessions. All were naive concerning psychophysical comparison experiments.

Stimuli and Design. A professional draftsman (using drafting tape and a magnifying glass) produced 20 pairs of black, horizontal line-length stimuli on white matte paper. The paired lines were made of 1/8-in. tape 1.59 mm wide, positioned horizontally side by side, and separated by 5-mm spaces. Four of the 20 pairs, used for practice, will not be considered further. The other 16 pairs were derived from the factorial combination of four levels of difference (4, 3, 2, and 1 mm) with four ratios of the longer to the shorter extent: 4:3, 8:7, 16:15, and 32:31. The line pairs were then photographed, and 35-mm slides were prepared with the shorter line appearing equally often on the right and on the left. The stimulus slides were projected onto a smooth white wall in a dimly illuminated room, and upon projection the lines appeared as a bright white against a dark background. The slide of the longest pair spanned a distance of approximately 1 m. Each subject viewed the stimuli from a distance of 1.8 m.

For purposes of analysis, the 16 pairs (the units are in millimeters) were grouped, from shortest to longest, into six stimulus-length levels: L1 = [(4,3), (8,6), (12,9), (16,12)], L2 = [(8,7), (16,14), (24,21), (32,28)], L3 = [(16,15), (32,30), (48,45), (32,31)], L4 = [(64,60), (64,62)], L5 = [(96,93)], L6 = [(128,124)]. Each subject received the 16 pairs and their mirror images (counterbalancing the left-right presentation order) with three sets of verbally presented instructions: "select the shorter," "select the longer," and "are the lines the same or different?"² The subject responded by pressing either the left or the right microswitch on a response panel, located just in front of the subject, who was seated at a table. The switch stopped a LaFayette electronic timer, which was started by a pulse from a light-sensitive switch mounted just in front of the lens of a Leitz Pradovit projector. The response button also advanced the projector and initiated the next trial. The clearly audible sounds of the advance of the projector served as the warning signal for each trial.

Each session involved 120 trials, with the task instructions fixed for two complete replications of the 20 pairs in the set. In every block of 20 trials, the first 4 pairs served as practice. Each of the six possible orders of task instruction was used four times, providing a total of 24 observations for each of the 20 pairs under each

of the three instructions. Thus, at the completion of the 12 sessions, each subject had made a total of 1,440 judgments. A different random ordering of pairs was used in every block of the set of 20 pairs.

Two subjects were instructed to sacrifice speed for accuracy; they were also told that they were competing with other subjects, and that the most accurate subject would receive an additional monetary reward. The remaining 2 subjects were instructed to sacrifice accuracy for speed by responding as quickly as possible, but not carelessly. The subjects were told that they would be informed regarding their accuracy at the conclusion of the experiment, and that trial-by-trial feedback would not be provided.

Results and Discussion

In Figure 1, mean response times under each instruction are plotted at each of the a priori defined stimulus-length levels, after averaging over the 24 replications with each pair and the 4 observers. A congruity effect is clearly evident. For the short line lengths, selection of the shorter line is faster than selection of the longer; conversely, for the relatively long lengths, selection of the longer line is faster. Interestingly, the crossover point is at the bisector of the stimulus range.

A four-way analysis of variance was performed on the mean overall response time for each of the three blocks of eight trials. Speed versus accuracy was the between-subjects factor and instruction (longer/shorter), stimulus length (six levels), and trial block were within-subject-factors. The main effect of stimulus-length level was significant [$F(5,10) = 15.42, p < .001$], confirming the a priori variation in difficulty. The interaction between stimulus-length level and instruction was reliable [$F(5,10) = 4.82, p < .017$], confirming the occurrence of semantic congruity effects for these comparisons. The interaction between speed-accuracy orientation and stimulus-length level was significant [$F(5,10) = 3.69, p < .037$],

Figure 1. Mean response times with each instruction, as a function of stimulus-length level in Experiment 1.

Figure 2. Semantic congruity indices plotted separately for each subject at each of the stimulus-length levels used in Experiment 1. The index is obtained by subtracting the mean overall response time with the instruction "shorter" from the mean overall response time with the instruction "longer" at each length level.

showing that the effects of increasing difficulty are amplified with accuracy-instructed subjects.

In Figure 2 an index of the semantic congruity effect is plotted for each subject, as a function of the stimulus-length level. The level is defined by the average length of the stimuli in the a priori classes defined earlier. Following Jamieson and Petrusic (1975c), the semantic congruity index is obtained by subtracting mean judgment times with the instruction "shorter" from the mean times with the instruction "longer"; the index when plotted at each level provides a convenient view of the instruction $\times$ stimulus pair interaction, which defines the semantic congruity effect. These plots show that the semantic congruity effect shown in Figure 1 is not an artifact of averaging over subjects. For each of the 4 subjects, the index is positive to the left of the figure and negative to the right, indicating faster "shorter" responses with the shorter lines and faster "longer" responses with the longer stimuli.

Table 1 shows that the speed-instructed subjects generally made more errors and responded more quickly than the accuracy-instructed subjects, which confirms the general effectiveness of the instructional manipulation of the relative demands for speed and accuracy. The plots in Figure 2 also suggest that the instructional manipulation of the relative demands for speed and accuracy are evident in the magnitude of the semantic congruity effect.

Table 1
Mean Response Time and Proportion Correct
for Each Subject in Experiment 1

	Speed		Accuracy	
	J.M.	J.J.	W.B.	D.K.
Judgment time (msec)	555	893	1280	1417
Proportion correct	.83	.92	.88	.94

Especially large effects can be observed for the accuracy-instructed subjects; the 900-msec effect seen here for subject W.B. is clearly an order of magnitude larger than those typically reported, perhaps because no previous study has employed a wide range of pairs varying in decision difficulty. On the other hand, the effect is extremely small for speed-instructed subjects, especially for subject J.M.; here, the effect is no more than 20 msec.

Examination of the plots in Figures 1 and 2 also reveals a dependence of the magnitude of the semantic congruity effect on the difficulty of the comparative judgment. Jamieson and Petrusic (1975c) showed that, with difficulty held constant, the magnitude of the semantic congruity effect increases systematically as extremity increases. Petrusic and Jamieson (1979) also showed response times to be exquisitely sensitive to both the ratio and the difference of the extents to be discriminated. Specifically, the closer the ratio of the longer to the shorter extent is, relative to an unbiased criterion of 1.0, the more difficult the comparison. In the present experiment, the pairs at Length Levels L1, L2, and L3, viewed a priori as relatively short, have the property that the ratios decrease (approach 1.0), and consequently the difficulty of discrimination monotonically increases, going from Level L1 to L3. Thus, for these relatively short pairs, the effects of increases in decision difficulty must overcome the naturally powerful effects of increases in extremity. Figures 1 and 2 show that for these relatively short pairs, the semantic congruity effect does increase, going from Level L1 to L3. On the other hand, for pairs at Length Levels L4, L5, and L6, viewed a priori as relatively long, the difficulty is lowest for L4 and remains constant thereafter (the ratio is 32:31 for pairs at Levels L5 and L6).³ Consequently, the mag-

nitude of the semantic congruity effect should increase going from pairs at Level L4 to pairs at the most extreme level, L6, because the pairs become more extreme and decision difficulty increases. Figures 1 and 2 show that this is the case.

Table 2 provides the proportion of correct judgments with each form of the comparison at each stimulus pair location, separately for each subject. The table shows that accuracy of discrimination did indeed vary with a priori difficulty: accuracy monotonically decreases over Length Levels L1 through L3 and decreases from L4 through L6. More importantly, it is clear that semantic congruity effects are not evident in the accuracy of response; the proportion of correct discriminations does not appear to depend in any systematic manner on the form of the comparison.

EXPERIMENT 2

In Experiment 1, the occurrence of semantic congruity effects was established with one type of perceptual comparison: line length. Experiment 2 was undertaken to establish the generality of the effect in perceptual comparisons using the heaviness continuum. Experiment 2 was also designed to test the accuracy of Henmon's (1911) introspective observation that confidence was also subject to semantic congruity effects. This extension of semantic congruity research to the confidence-judgment measure is also important because it permits further examination of the basis for the judgment of confidence. For example, according to Vickers's (1979) balance-of-evidence view (see also Smith & Vickers, 1988), since both response accuracy and confidence are determined by the balance of evidence (the difference in the accumulated evidence totals favoring each response), both accuracy and confidence must simultaneously exhibit or fail to exhibit the semantic congruity effect. Finally, Experiment 2 was designed to examine further the surprising dissociation, observed in Experiment 1, between the accuracy-based and response-time-based measures of the semantic congruity effect by continuing to employ stimulus pairs imperfectly discriminated with a substantial error rate.

Method

Subjects. Eight university students served for six 1-h sessions for payment of \$12 and course credit. None had previously participated in a psychological experiment.

Apparatus. Small plastic bottles, filled with lead shot and held securely in place with paraffin, were used to define the required weights. Prior to lifting each weight, the subject's hand rested on a metal disk. A switch directly below the disk was closed, starting a Hewlett-Packard counter, whenever the subject lifted a weight. The counter operated during the lifting of the first stimulus; it was reset manually before the presentation of the second stimulus and again after the response, which was made by pressing either of two buttons on a response box. Each of the response buttons was connected to a light; the onset of either light indicated to the experimenter which response button had been pressed. The screen that was used to conceal the disk, the weights, and the experimenter from the subject consisted of a plywood box with an opening mea-

Table 2
Mean Proportion Correct for Each Subject for Each
Stimulus-Length Level with Each Instruction in Experiment 1

Subject	Instruction	Length Level					
		L1	L2	L3	L4	L5	L6
J.M.	Select shorter	.98	.81	.69	.79	.63	.75
	Select longer	1.00	.81	.71	.81	.75	.71
J.J.	Select shorter	1.00	.95	.86	.88	.92	.75
	Select longer	1.00	.99	.83	.88	.96	.75
W.B.	Select shorter	1.00	.89	.73	.85	.96	.58
	Select longer	1.00	.99	.73	.83	.92	.67
D.K.	Select shorter	.98	.94	.84	.98	.92	1.00
	Select longer	1.00	.96	.80	.98	.98	.98
Mean	Select shorter	.99	.90	.78	.88	.86	.77
	Select longer	1.00	.94	.77	.88	.90	.78

suring 47×16 cm, through which the subject's hand extended. The surface of the box sloped upward, so that the opening on the experimenter's side measured 47×41 cm. A Uher tape recorder was used for recording and playback of the instructions for lifting and lowering the hand.

Stimuli. The stimulus set comprised 23 weights derived from the factorial combination of the ratios 2:3, 4:5, and 10:11 with the three differences 10, 25, and 50 g: 9 consisted of the pairs (20,30), (50,75), (100,150), (40,50), (100,125), (200,250), (100,110), (250,275), (500,550); 9 included their mirror images (30,20), ..., (550,500); the other 5 were the identical pairs (100,100), (200,200), (300,300), (400,400), (500,500). The first component in each ordered pair indicates the first presented stimulus. On each block of 46 trials, each of the 23 stimulus pairs was presented once with the instruction to select the lighter member of the pair and once with the instruction to select the heavier. Each subject received two such blocks of trials in every experimental session. The first session was viewed as practice and was analyzed separately. Over the remaining 5 sessions, 10 replications were obtained for each pair, for a total of 460 observations per subject. For purposes of analysis and to permit greater stability of estimates of average response times, confidence ratings, and judgmental accuracy, the pairs were grouped into four weight levels with each instruction at each location on the continuum: $W1 = [(20,25), (25,20), (40,50), (50,40), (50,75), (75,50)]$; $W2 = [(100,100), (100,110), (110,100), (125,150), (150,125), (150,175), (175,150)]$; $W3 = [(200,200), (200,250), (250,200), (250,275), (275,250), (300,300)]$; $W4 = [(400,400), (500,500), (500,550), (550,500)]$.

Procedure. Detailed instructions were read to the subject, indicating the placement of the hand, palm upwards, on the metal disk, the coordination of the sequence of lifting and lowering with the tape-recorded playback of the recorded instruction sequences, and the use of the response buttons. The sequence of events on each trial, controlled by the tape-recorded instructions, consisted of the word "ready," followed by the instruction ("lighter" or "heavier"), placement of the first weight in the subject's hand, presentation of the instruction "lift," which was followed 2 sec later by the instruction "lower," the removal of the weight from the subject's hand, a 450-msec interstimulus interval, the placement of the second weight in the subject's hand, and, finally, the instruction to "lift." The subject then indicated the response by pressing, with the index finger of the left hand, the button marked "first" or the button marked "second," for the first or second weight. In addition to making this comparative judgment, the subject was instructed to report the instruction for each trial orally and to give a confidence rating for the comparative judgment. A 6-point scale was used, with 0 indicating a guess and 5 complete confidence or certainty. All responses and decision times were recorded by the experimenter. Presentation of the ready signal began 5 sec after the instruction to lift the second weight, effectively forcing a 5-sec response deadline. The instructions also indicated that both speed and accuracy were equally important. As in the first experiment, trial-by-trial feedback was not provided.

Results and Discussion

For both comparative judgment times and confidence ratings, means were obtained for each subject, for each instruction, and for each stimulus set, by averaging over trials and over pairs within the a priori stimulus sets that defined the various locations on the weight continuum. These means served as the dependent variables throughout subsequent analyses. The proportion of correct judgments, averaged over trials and subjects, served as the accuracy measure.

Response-time analyses. The upper portion of Figure 3 provides plots of mean response times with each instruc-

tion at each of the four a priori weight levels after averaging over all subjects and the 20 replications with each instruction for each of the pairs at the weight level. As in Experiment 1, the level is defined by the average of the weights in the a priori class. In replication of Experiment 1, the plots in Figure 3 provide clear evidence of a semantic congruity effect.

A two-way analysis of variance with weight level and response instruction as within-subject factors showed the interaction of these two factors to be significant [$F(3,21) = 6.27, p < .003$], thereby confirming the reliability of the semantic congruity effect for these heaviness comparisons. The weight levels also differed with respect to overall response time [$F(3,21) = 11.27, p < .001$]. From lightest to heaviest, the overall mean response times were 1,283, 1,309, 1,424, and 1,540 msec, respectively. As in the first experiment, the main effect of instruction failed to attain significance [$F(1,7) = 1.85, p > .21$], indicating an absence of lexical marking for the response-time measure.

Confidence-rating analyses. The lower panel of Figure 3 provides plots of the average confidence rating with each instruction for each of the weight levels. Paralleling the plots in the upper panels for the response-time measure, a crossover effect is clearly evident for the

Figure 3. Mean response times (upper panel) and mean confidence ratings (lower panel) at each of the a priori stimulus-weight levels with each instruction in Experiment 2.

confidence-rating measure: for the relatively light pairs, confidence is higher when the lighter member in the pair is selected, and for the relatively heavy pairs, confidence is greater when the heavier member of the pair is selected.

Taken together, the plots in the upper and the lower portions of Figure 3, in addition to demonstrating the joint occurrence of the crossover effect, show that the two measures are, as is well known, inversely related (see, e.g., Henmon, 1911; Vickers & Packer, 1982). The plots also show that the crossover point occurs at nearly the same point—between the 100- and the 200-g stimulus pairs.

An analysis of variance, paralleling that for response time, showed only the interaction between weight level and response instructions to be statistically significant [$F(3,21) = 10.15, p < .001$], confirming the reliability of the semantic congruity effect with the confidence-rating measure.

Early congruity effects. In Figure 4, mean response times with each instruction are plotted as a function of weight level for the combined data of all subjects, separately for the first and second blocks of the practice session.⁴ Contrary to the faster responding with the instruction "heavier," expected on the basis of the lexical markedness notion (see Clark, 1969), the plots for Block 1 show a funnel form of semantic congruity effect. For the lightest of these relatively light pairs, responses are faster with the instruction to select the lighter and responses are equally fast with the two instructions for the heaviest of

Figure 4. Mean response time and confidence report with each instruction, plotted as a function of weight level, for the first and second block of practice trials in Session 1 of Experiment 2.

Table 3
Mean Proportion Correct for Each Stimulus-Weight Level with Each Instruction in Experiment 2*

Instruction	Stimulus Weight Level			
	W1	W2	W3	W4
"Lighter"	.724	.761	.686	.605
"Heavier"	.741	.774	.721	.638

*Proportion based on 480, 560, 480, and 320 observations for the four respective stimulus-weight levels, after averaging over pairs within sets and the 8 subjects. The average weight of the stimulus pairs in the locations defined by weight levels W1, W2, W3, and W4, is 43.3, 128.2, 245.8, and 487.5 g, respectively.

the stimuli (those at Level 4). On the other hand, the plots for Block 2 show that the full semantic congruity effect is evident for the second block of practice trials.

Analyses of variance with the response-time measure showed the main effect of instruction to be significant [$F(1,6) = 9.76, p < .021$], and the interaction between instruction and weight level was also significant [$F(3,18) = 3.79, p < .028$]. With the confidence-rating measure, only the weight level $\times$ instruction interaction proved reliable [$F(3,18) = 4.32, p < .018$], attesting to the occurrence of the semantic congruity effect at the outset of the experiment.

Response-accuracy analyses.⁵ Table 3 presents the mean proportion of correct judgments with each instruction for each of the weight levels, after averaging over subjects and pairs within each stimulus set. Slightly, although consistently, higher accuracy with the instruction to select the heavier than with the instruction to select the lighter is evident, suggesting the presence of a lexical markedness effect. However, as in the first experiment, response accuracy fails to reveal a semantic congruity effect, unlike the response-time or confidence-rating measures.

In summary, Experiment 2 provides further evidence of reliable semantic congruity effects with yet another perceptual continuum—the heaviness of successively lifted weights. In confirmation of Henmon's (1911) introspective analyses of psychophysical comparisons, Experiment 2 extends the range of the semantic congruity effect to include confidence judgments. In replication of the first experiment, semantic congruity effects, although evident with response time, do not occur with the response-accuracy measure.

EXPERIMENT 3

As discussed earlier, studies of perceptual semantic congruity have, to date, employed nonconfusable stimuli (errors accounting for 1%–5% of the data). Moreover, only correct response times have typically been analyzed. This is unfortunate, because the analysis of error response times and, more specifically, of the semantic congruity effect for error response times provides an important test of alternative theories of the semantic congruity effect. Specifically, according to the semantic coding view, as previously stated, stimulus-code translation occurs after the

discrimination is performed. Although the discrimination time for errors is typically longer than for correct responses when conditions emphasize accuracy (see, e.g., Petrusic & Jamieson, 1989; Pike, 1971), the code translation component, which accounts for the congruity effect, is supposedly a constant time increment. Moreover, since errors can also occur as a consequence of the failure to translate codes and thereby reduce the semantic congruity effect's magnitude, the overall congruity effect for error responses cannot be larger than for correct responses.

To test this prediction of the semantic coding model we used stimulus difficulty ratios that, on the basis of previous research (Petrusic & Baranski, 1988b), were known to provide many errors. Also, in order to provide further tests of the generality of the congruity effect with purely perceptual stimuli, we utilized yet another continuum: horizontal visual proximity.

Method

Subjects. Twenty Carleton University undergraduate students with normal or corrected-to-normal vision participated in one approximately 90-min experimental session in order to satisfy course requirements. All the subjects were naive with respect to the nature and aims of the experiment.

Apparatus. The stimuli and instructions were presented on an Amdek 310A video monitor controlled by an IBM-XT-compatible computer that also controlled trial presentation, event sequencing, randomization, and response recordings. Responding was performed on an IBM PC mouse positioned beside the keyboard, on the side of the subject's preferred hand. The keyboard was used in the reporting of confidence. A small desk lamp behind the video monitor provided sufficient light for responding but did not interfere with the visual display.

Stimuli. On each trial, the visual display consisted of two 5-mm vertical lines presented on the left and right of a central fixation marker (10-mm vertical line). All lines were 1 mm (3.13 pixels) wide; they were controlled for brightness and contrast at the outset of each session.

The central line served the purposes of allowing a referent from which to perform the comparative judgment and to divide the video screen (which has a horizontal resolution capability of 720 pixels) into a left and right side. In addition, the central marker was presented equally often either bisecting the screen or 2 pixels to the left or right of the center of the screen. The three levels of this nuisance factor ensured that the subjects performed the comparison in relation to the referent and did not rely on estimates of the distance from the edge of the video monitor.

The notation (x, y) will be used to describe a stimulus pair x pixels to the left of the center line and y pixels to the right. Twelve stimulus pairs were used in the experiment, expressed in pixels: (30,28), (28,30); (50,48), (48,50); (70,69), (69,70); (260,243), (243,260); (280,269), (269,280); and (300,296), (296,300). Upon collapsing over the two pairs within each distance level [i.e., over (x, y) and (y, x)], the pairs, for convenience of identification, will be referred to as Pairs 1 through 6, respectively.

As previously stated, Petrusic and Jamieson (1979) have shown that the difficulty of comparative judgments can be effectively manipulated by means of variation of the ratio of the longer to the shorter extent (x/y) of the comparison pair. Thus, the larger the ratio (relative to the criterion of 1.0), the easier the comparison. Here, the stimulus pairs are expressed, in terms of difficulty, by the ratios 1.07, 1.04, 1.01, 1.07, 1.04, and 1.01 for Pairs 1 through 6.

Procedure. The subjects were seated approximately 60 cm from the video screen in a dimly lit room. The stimuli were presented at eye level, and the nearest and farthest stimulus pairs subtended a visual angle of approximately 3° and 14°.

At the outset of the session, the subjects were read the following instructions:

On the video monitor you will be presented stimuli, in the form of small vertical lines, on the left and right of a central marker (represented by a slightly longer vertical line). Your task is to respond to the instructions "NEARER" or "FARTHER" by pressing the appropriate response button (left or right) on the "mouse," indicating which small vertical line is respectively nearer or farther from the central line. Please do not use the edge of the screen as an aid for your judgment, as, the central line will be varied in its position from trial to trial. We ask you to be as accurate as you can while not taking too much time to respond. Thus, for the purposes of this experiment, both speed and accuracy are equally important. In addition, after each response, you will be prompted to evaluate your accuracy by providing a confidence report. Do this by entering a confidence rating (using the numbered keys at the top of the keyboard before you) from 0 to 100. A rating of 100 would indicate complete certainty in your response, 50 a guess, and 0 complete certainty in an error. All intermediate ratings should, accordingly, relate to varying degrees of certainty or uncertainty in the accuracy of your response. There is no time limit on registering your confidence. After entering your confidence simply strike the 'Enter' key to initiate the next trial. There will be several break periods during the session and when you are ready to continue, strike any key to commence the next block of trials.

Each session began with 24 practice trials, followed by four blocks of 96 trials, each of which was separated by an instructed break period.⁶ All pairs were counterbalanced for the left-right presentation order and randomized within each block. Furthermore, the two instructions appeared equally often within each block and with each stimulus pair.

Each trial began with the presentation of an instruction ("NEARER" or "FARTHER"), followed 1,000 msec later by the stimulus pair. The instruction and pair remained on the screen until completion of the comparative judgment. Following the response, the screen was cleared, and the subjects were prompted to enter their confidence rating.

The subjects were randomly assigned to one of two groups. The groups were identical, except that one received trial-by-trial response feedback ("Correct" or "Incorrect") following each confidence report and preceding the next trial, whereas the other did not.

Results and Discussion

The data for one subject, who provided a mean response time under 200 msec (51% correct overall), was excluded. Another subject was run as a replacement. Ten trials (0.14% of the overall 7,904 trials) with response times either faster than 200 msec or longer than 10 sec were discarded.

Separate analyses of variance, with the mean response time, proportion correct, and probability confidence (arcsine transformed) as dependent measures, were used to analyze the data. Feedback versus no feedback condition was a between-groups factor, and stimulus pair (with six levels) and instruction ("NEARER" vs. "FARTHER") were within-subject factors. Since there were neither significant main effects nor interactions involving feedback, we present the data collapsed over all 20 subjects.

The results are presented in two main sections. The first presents the global crossover effects for the three dependent measures. The second presents the crossover effects

separately, for the correct and the error responses for the response-time and the confidence-rating measures.

Global effects. In Figure 5, the mean response time, probability confidence, and proportion correct for each instruction are plotted as a function of visual extent level. The plot provides several points of interest. First, from a global perspective, the response times are fastest, and accuracy and confidence highest, for the most discriminable pairs (1,4). The response times increase and accuracy and confidence decrease for the intermediate-level-difficulty pairs (2,5). Finally, response times are longest, and accuracy and confidence lowest, for the most difficult pairs (3,6). These results are confirmed by a significant main effect of stimulus pair for both response time [$F(5,90) = 6.58, p < .002$] and response accuracy [$F(5,90) = 48.55, p < .0001$].

Second, the plot clearly shows a large crossover effect for both the response-time and the confidence-rating measures, but not for response accuracy. The results are supported by a highly significant pair $\times$ instruction interaction for response time [$F(5,90) = 12.86, p < .0001$]

Figure 5. Mean response time, probability confidence, and proportion correct for each instruction, plotted as a function of visual extent level in Experiment 3.

Figure 6. Semantic congruity magnitude for response time (top panel) and confidence (lower panel) for error (filled circles) and correct (open circles) responses, plotted as a function of visual extent level in Experiment 3.

and confidence [$F(5,90) = 9.23, p < .0001$], but not for response accuracy [$F(5,90) = 2.14, p > .05$].

In the main, the plots provided in Figure 5 provide a replication of the dependence of semantic congruity magnitude on the difficulty of comparison that is evident in Experiment 1. For the three near pairs, semantic congruity effects are of approximately equal magnitude; we would expect the smallest congruity effect for Pair 3 if difficulty exerted no effect. On the other hand, for the three far pairs, since the pairs become both more difficult and farther (i.e., more extreme) on the continuum simultaneously, semantic congruity effects monotonically increase with extremity.

Error versus correct congruity effects. In Figure 6, the magnitude of the semantic congruity effect is plotted separately for correct and error responses, for both response time (top panel) and confidence (lower panel). The results are striking. Error response semantic congruity is larger than that for correct responses for each of the six stimulus pairs for both the response-time and the confidence measures. Simultaneously, although not shown, error response times are longer than correct response times and confidence associated with error

responses is lower than that with correct responses for each of the six stimulus pairs.

GENERAL DISCUSSION

These experiments establish that semantic congruity effects can be obtained with perceptual comparisons, and, at the same time, the experiments provide the limiting conditions for such occurrences. In replication of the earlier studies using supraliminally different stimuli (Banks et al., 1982; Marschark & Paivio, 1981; Petrusic & Baranski, 1988a), when performance is errorless, as in Experiment 1 with the shortest line pairs, the effect is minimal. However, when performance is not errorless, as a consequence of decreases in discriminability, the effect is then readily obtained. Moreover, these findings are based on three different perceptual continua not previously examined for semantic congruity effects: classical comparisons of simultaneously presented line lengths, of successively presented weights, and of visual distances. As a consequence, these findings also establish the generality of the effect with confusable perceptual magnitudes.

As well as establishing the conditions for the full occurrence of the effect with perceptual comparisons, the present findings permit a number of theoretical points. Banks et al. (1982) have taken the failures to obtain the semantic congruity effect with perceptual comparisons as evidence in support of their theoretical position. Actually, Banks et al. (1982) have taken the extreme position that congruity effects occur with perceptual comparisons only if some form of symbolic interpretation of the stimuli occurs. Thus, according to the semantic coding position, since semantic congruity effects are not clearly evident with errorless perceptual comparisons, these comparisons must involve "purely perceptual" processes. However, since the semantic congruity effect is obtained with confusable perceptual stimuli, these comparisons must not involve "purely perceptual" processes. Thus, according to Banks et al. (1982), these "stimuli are interpreted as if they were symbols denoting positions on a unidimensional semantic continuum" (p. 276). However, it is not readily apparent why whether processing is purely perceptual or not should depend on whether discrimination is errorless or not. Alternatively, perhaps semantic coding is involved in all cases of perceptual comparative judgments, and thus the semantic congruity effect occurs with both confusable and nonconfusable stimuli, although these effects are too small to be detected when supraliminally different perceptual comparisons are involved.⁷

At a general level, the occurrence of semantic congruity effects with evidently all classes of comparative judgments removes an important set of procedures for inferring differences between processing memorial and processing sensory (perceptual) stimulus information (cf. Banks et al., 1982). Nevertheless, it should be noted that the effect with remembered comparisons is readily obtained even when performance is errorless, contrary to the perceptual comparison case.

The question of whether some form of verbal coding is involved in perceptual comparisons is difficult and important, but it cannot be decided with the present data. Indeed, it is doubtful that perceptual comparisons are conducted without resort to a number of memory processes, even when stimuli are presented simultaneously for comparison. The critical question of whether this use of memory takes the form postulated by the semantic coding theory remains. In the absence of direct tests for the presence or absence of verbal semantic coding, the role of semantic coding in perceptual comparisons can be examined indirectly only by asking how well the semantic coding theory is able to account for the present findings.

The findings of the present studies imply a decisional locus for the congruity effect, which is contrary to the semantic coding theory. The substantial dependence of the magnitude of the effect on the relative demands for speed and accuracy found in Experiment 1 is clearly not permitted by the semantic coding theory, not at least as it is currently formulated (see Banks, 1977; Banks, et al., 1976). Of course, post hoc, the semantic coding theory would require that the probability of code translation be dependent on the relative demands for speed and accuracy; when subjects are severely speeded, code translation would not occur on some proportion of trials, and the semantic congruity effect would thereby be reduced. However, it must then be the case that semantic congruity effects would be especially evident with the response-accuracy measure (i.e., in terms of the probability of correct responses), when responses are speeded at the expense of accuracy. As noted earlier, semantic congruity effects are not evident with the accuracy measure even under speeded conditions.

Furthermore, the semantic coding theory is unable to predict the particular form of dependence of semantic congruity magnitude on decision difficulty obtained in Experiments 1 and 3, for example. Notably, for pairs (x_1, y_1) and (x_2, y_2) , whenever $x_1 < y_1 < x_2 < y_2$ and $(y_1 - x_1) > (y_2 - x_2)$, and whenever responses with the instruction "select the shorter" are faster than with the instruction "select the longer," we can infer that each of the elements in these pairs is very likely to be coded as in the "short" category (i.e., given the semantic congruity effect). In fact, a priori, and surely on the basis of an overt binary categorization task, the elements in each of these pairs would be in the "short" category. According to the semantic coding theory, since both members of the pair (x_1, y_1) are more extreme on the continuum than elements of the pair (x_2, y_2) , they are both more likely to be coded in the same category (see Banks et al., 1976); consequently, the pair (x_1, y_1) is more likely to require code translation than is the pair (x_2, y_2) , and therefore the pair (x_1, y_1) should always exhibit the larger semantic congruity effect. However, the findings from Experiment 1 (and 3) show that the opposite result is obtained with this difficulty manipulation. In each case, the larger semantic congruity effect is obtained with the less extreme stimulus pair because it is less discriminable.

The fact that the crossover effect is larger for error than correct response times also adds further support for a decisional locus. Here as well, the magnitude of the congruity effect mirrors base-rate decisional response times precisely. On average, with each of the six stimulus pairs, response times are longer and confidences lower with error responses than with correct responses, and semantic congruity magnitude is larger for both response times and confidence estimates with error responses.

Several theories of the comparison process admit a decisional locus for the semantic congruity effect; the present findings provide general support for this class of theories. Audley and Wallis (1964), in addition to demonstrating the effect with brightness comparisons, provided a fully developed quantitative theory of the semantic congruity effect. According to this theory, as in all evidence-accrual theories, the overt response is made only after one of the competing response systems accrues a criterion amount of partial information or evidence. Specifically, Audley and Wallis assumed that a run of implicit responses of one type before the other triggered the overt response, following the discrete stochastic counter model developed originally by Audley (1960). The key idea accounting for the semantic congruity effect is that the rate of evidence accrual with a given instruction is directly dependent upon the location of the stimuli on the attribute. For example, when two relatively heavy alternatives are presented, information regarding which one is heavier accrues more quickly than information regarding which one is lighter. The semantic congruity effect then arises because the overt response would be delayed for the instruction to choose the lighter, since more time would be required for the criterion run of implicit "light" responses. Conversely, when the stimuli are both relatively light, the required run of implicit "light" responses will be obtained sooner than the run of "heavy" responses. One very striking prediction of this model is that a reverse semantic congruity effect (i.e., higher accuracy with the instruction to select the lighter when the alternatives are both heavy) will occur with the response-accuracy measure (see Audley & Wallis, 1964, p. 72). This specific model is thus disconfirmed, because semantic congruity effects simply are not evident with the accuracy measure.

Jamieson and Petrusic (1975c) and Marks (1972) developed alternative versions of a theory which postulated that presentation of a particular instruction (e.g., "select the heavier") resulted in the activation of an ideal point on the underlying continuum, in the Coombs (1950, 1964) unfolding theory sense. According to Jamieson and Petrusic (1975c), the overt judgment is based on a comparison of the ratio of the distances of each stimulus from the reference point activated by the presented instruction relative to a decisional criterion (the criterion ratio is 1.0 when it is unbiased, and a biased criterion captures the notion of lexical markedness). This ratio is generally nearer to the criterion when the ideal point activated by the instruction is far from the representations of the

presented stimuli than it is when the ideal point is close to the stimulus representations. This fact, together with the assumption that response time is a monotonically decreasing function of the difference between the ratio and the criterion, results in the semantic congruity effect. According to Marks (1972), the precision of the representation of the stimuli is inversely dependent on the distance of the locations of the stimuli from the activated ideal point. Thus, in both models, the semantic congruity effect arises because of a deterioration of the quality of the information on which the overt decision is based.

Holyoak (1978), using the method of triads in numerical magnitude comparisons, extended the discrepancy ratio notion developed in Jamieson and Petrusic (1975c) by embedding it in a specific model (the simple random walk) of the accumulation of evidence relative to a decisional criterion. Acknowledging the broader dependence of judgmental choice behavior on contextual factors (cf. Helson, 1964), he aptly called this hybrid model, which included the discrepancy ratio and the random walk, a *reference-point* theory of the comparison process. In its current form, the Holyoak (1978) reference-point model does not provide a full account of the present findings—especially those in Experiment 3, which show uniformly longer response times for error responses. The simple random walk model predicts equal times for correct and error responses (see Laming, 1968; Luce, 1986; Townsend & Ashby, 1983). Consequently, a model of the evidence-accrual process other than the simple random walk model must be implemented in the context of the reference-point theory (e.g., see Ashby, 1983; Link & Heath, 1975; Petrusic & Jamieson, 1978; Smith & Vickers, 1988; Vickers, 1979).

The pattern of semantic congruity effects observed over the two blocks of the initial session of practice trials in Experiment 2 provides information concerning the dynamic regulation of reference-point location. Let I^* and I_* denote the reference points activated upon the initial presentation of instructions to select the heavier and the lighter of the two weights, respectively. Let *Max* and *Min* denote the estimates of the largest and the smallest stimuli encountered in, say, a given block of trials. The reference-point theory suggests that the full crossover effect will not be found whenever the bisector of the reference points is greater than *Max* [i.e., when $(I_* + I^*)/2 > \text{Max}$]; rather, a funnel effect will be obtained, and the neck of the funnel will occur with the heaviest of the weights encountered. However, since by the second block of practice trials the full semantic congruity effect is present, we can infer a very rapid shift of the reference point I^* toward that of the value of *Max*. This pattern of funnel and semantic congruity effects at the very outset of the comparison process suggests that the reference point used at any time is a compromise between values stored in long-term memory, reflecting tacit knowledge about the dynamic ranges of a variety of sensory continua, and the perceived within-session range. This compromise might well be

described by the simple equation $aI_* + (1-a)\text{Min}$ and $bI^* + (1-b)\text{Max}$, where a and b are weights varying between 1 and 0.

These applications of reference-point-theory ideas reinforce the view that various memory processes are continually involved in perceptual comparisons. At the least, retrieval of reference points from long-term memory is required, as are estimates of the maximum and minimum magnitudes encountered in the session, and these must be retained and continuously updated. Of course, the particular view of the memory processes involved depends on a theoretical position concerning the semantic congruity effect. The semantic coding theory could also provide an equally plausible view of the practice-session data from Experiment 2, in terms of the joint dependence of long-term knowledge about the heaviness continuum and dynamic regulation of semantic codes. However, in the balance, the evidence from the current data tends to favor the reference-point view over the semantic coding position.

Extension of the semantic congruity effect to include the confidence-rating measure but not the response-accuracy measure has also proved theoretically revealing. First, the simultaneous occurrence of the semantic congruity effect with the response-time and confidence-rating measures and not the response-accuracy measure is difficult to understand in light of Vickers's (1979; see also Vickers & Packer, 1982) balance-of-evidence notion (difference in accumulated evidence for each response) in the context of his accumulator model. According to this view, whenever the balance of evidence is large, confidence will be high, as will accuracy. However, it is well known that the three response measures differ considerably in sensitivity. For example, Petrusic and Jamieson (1979) showed that the response-time measure continues to be sensitive to arithmetic relations defined on stimulus pairs well after the conventional discriminative accuracy measure (e.g., d' or proportion correct) becomes insensitive to these variations. This dissociation between response time and probability of response is not without precedent. For example, Greenberg and Doren (1965), in their tests of alternative interpretations of the unfolding process in Coombs' (1950, 1964) theory of unidimensional preference, found clear variations in response time as the ideal point was systematically shifted, and no effect of these ideal point variations on response probability.

Similarly, the failure to obtain a congruity effect for the response-accuracy measure causes problems for the early versions of reference-point models (e.g., Holyoak, 1978; Jamieson & Petrusic, 1975c; Marks, 1972). In these models, response time and accuracy are jointly monotonic functions of the difference between the decisional criterion and the discrepancy ratio. However, the measures need not be equally affected by the distance of the discrepancy ratio from the decisional criterion. Indeed, it is possible that subjects are able to set a flexible decisional criterion in order to maintain an approximately constant accuracy level with each instruction. With this provision,

the reference-point models provide a qualitative account of the present findings.

REFERENCES

- ASHBY, G. (1983). A biased random walk model of two choice reaction times. *Journal of Mathematical Psychology*, 27, 277-297.
- AUDLEY, R. J. (1960). A stochastic model for individual choice behaviour. *Psychological Review*, 67, 1-15.
- AUDLEY, R. J., & WALLIS, C. P. (1964). Response instructions and the speed of relative judgments: I. Some experiments on brightness discrimination. *British Journal of Psychology*, 55, 59-73.
- BANKS, W. P. (1977). Encoding and processing of symbolic information in comparative judgments. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 11, pp. 101-159). New York: Academic Press.
- BANKS, W. P., CLARK, H. H., & LUCY, P. (1975). The locus of the semantic congruity effect in comparative judgments. *Journal of Experimental Psychology: Human Perception & Performance*, 1, 35-47.
- BANKS, W. P., & FLORA, J. (1977). Semantic and perceptual processes in symbolic comparisons. *Journal of Experimental Psychology: Human Perception & Performance*, 3, 278-290.
- BANKS, W. P., FUJII, M. S., & KAYRA-STUART, F. (1976). Semantic congruity effects in comparative judgments of magnitudes of digits. *Journal of Experimental Psychology: Human Perception & Performance*, 2, 435-447.
- BANKS, W. P., MERMELSTEIN, R., & YU, H. K. (1982). Discriminations among perceptual and symbolic stimuli. *Memory & Cognition*, 10, 265-278.
- BANKS, W. P., & ROOT, M. (1979). Semantic congruity effects in judgments of loudness. *Perception & Psychophysics*, 26, 133-142.
- BANKS, W. P., WHITE, H., STURGILL, W., & MERMELSTEIN, R. (1983). Semantic congruity and expectancy in symbolic judgments. *Journal of Experimental Psychology: Human Perception & Performance*, 9, 560-582.
- CLARK, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review*, 76, 387-404.
- COOMBS, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, 57, 145-158.
- COOMBS, C. H. (1964). *A theory of data*. New York: Wiley.
- ELLIS, S. (1972). Interaction of encoding and retrieval in relative age judgments: An extension of the "crossover" effect. *Journal of Experimental Psychology*, 94, 291-294.
- FECHNER, G. T. (1966). *Elements of psychophysics* (H. E. Adler, Trans.). New York: Harper & Row. (Original work published 1860)
- FRIEND, K. E. (1973). Perceptual encoding in comparative judgments of face. *Memory & Cognition*, 1, 80-84.
- GREENBERG, M. H., & DOREN, D. G. (1965). Relationship between latency and remoteness in preference judgments. *Journal of Experimental Psychology*, 70, 182-188.
- HELLSTRÖM, A. (1977). Time errors are perceptual: An experimental investigation of duration and a quantitative successive-comparison model. *Psychological Research*, 39, 345-388.
- HELLSTRÖM, A. (1985). The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin*, 97, 35-61.
- HELVSON, H. (1964). *Adaptation-level theory: An experimental and systematic approach to behavior*. New York: Harper & Row.
- HENMON, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychological Review*, 18, 186-201.
- HOLYOAK, K. J. (1978). Comparative judgments with numerical reference points. *Cognitive Psychology*, 10, 203-243.
- HOLYOAK, K. J., & MAH, W. A. (1981). Semantic congruity in symbolic comparisons: Evidence against an expectancy hypothesis. *Memory & Cognition*, 9, 197-204.
- HOLYOAK, K. J., & PATTERSON, K. K. (1981). A positional discriminability model of linear-order judgments. *Journal of Experimental Psychology: Human Perception & Performance*, 7, 1283-1302.
- HOLYOAK, K. J., & WALKER, J. H. (1976). Subjective magnitude information in semantic orderings. *Journal of Verbal Learning & Verbal Behavior*, 15, 287-289.

- HOWARD, R. (1983). The semantic congruity effect: Some tests of the expectancy hypothesis. *Acta Psychologica*, **53**, 205-216.
- HOWELL, D. C. (1982). *Statistical methods for psychology*. Boston: Duxbury Press.
- JAMIESON, D. G., & PETRUSIC, W. M. (1975a). Pairing effects and time-order errors in duration discrimination. *Perception & Psychophysics*, **18**, 107-113.
- JAMIESON, D. G., & PETRUSIC, W. M. (1975b). Presentation order effects in duration discrimination. *Perception & Psychophysics*, **17**, 197-202.
- JAMIESON, D. G., & PETRUSIC, W. M. (1975c). Relational judgments with remembered stimuli. *Perception & Psychophysics*, **18**, 373-378.
- JOHN, I. D. (1975). A common mechanism mediating the time-order error and the cross-over effect in comparative judgments of loudness. *Australian Journal of Psychology*, **27**, 51-60.
- KOSSLYN, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- KOSSLYN, S. M., MURPHY, G. L., BEMESDERFER, M. E., & FEINSTEIN, K. J. (1977). Category and continuum in mental comparisons. *Journal of Experimental Psychology: General*, **4**, 341-375.
- LAMING, D. (1968). *Information theory of choice reaction times*. New York: Academic Press.
- LINK, S. W., & HEATH, R. A. (1975). A sequential theory of psychological discrimination. *Journal of Mathematical Psychology*, **40**, 77-105.
- LUCE, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- MARKS, D. F. (1972). Relative judgment: A phenomenon and a theory. *Perception & Psychophysics*, **11**, 156-160.
- MARSCHARK, M., & PAIVIO, A. (1979). Semantic congruity and lexical marking in symbolic comparisons: An expectancy hypothesis. *Memory & Cognition*, **7**, 175-184.
- MARSCHARK, M., & PAIVIO, A. (1981). Congruity and the perceptual comparison task. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 290-308.
- MOYER, R. S. (1973). Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics*, **13**, 180-184.
- MOYER, R. S., & DUMAIS, S. T. (1978). Mental comparison. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 12, pp. 117-155). New York: Academic Press.
- MUNSTERBERG, H. (1894). Studies from the Harvard psychological laboratory: A psychometric investigation of the psychophysical law. *Psychological Review*, **1**, 45-51.
- PETRUSIC, W. M., & BARANSKI, J. V. (1988a). *Binary and quaternary relational judgments with remembered magnitudes: Extending the internal psychophysics*. Manuscript in preparation.
- PETRUSIC, W. M., & BARANSKI, J. V. (1988b, August). *Context, context shifts, and semantic congruity effects in comparative judgments*. Paper presented at the XXIV International Congress of Psychology, Sydney, Australia.
- PETRUSIC, W. M., & JAMIESON, D. G. (1978). Relation between probability of preferential choice and time to choose changes with practice. *Journal of Experimental Psychology: Human Perception & Performance*, **4**, 471-482.
- PETRUSIC, W. M., & JAMIESON, D. G. (1979). Resolution time and the coding of arithmetic relations on supraliminally different visual extents. *Journal of Mathematical Psychology*, **19**, 89-107.
- PETRUSIC, W. M., & JAMIESON, D. G. (1989). Comparative judgments of the ease of sameness-difference judgment: Matching probabilistic structures. *American Journal of Psychology*, **102**, 69-90.
- PIKE, A. R. (1971). The latencies of correct and incorrect responses in discrimination and detection tasks: Their interpretation in terms of a model based on simple counting. *Perception & Psychophysics*, **9**, 455-460.
- SCHAB, F. R., & CROWDER, R. G. (1988). The role of succession in temporal cognition: Is the time-order error a recency effect of memory? *Perception & Psychophysics*, **44**, 233-242.
- SHIPLEY, W. C., NORRIS, E. D., & ROBERTS, M. L. (1946). The effect of changed polarity of set on decision time of affective judgments. *Journal of Experimental Psychology*, **36**, 237-243.
- SMITH, P. L., & VICKERS, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, **32**, 135-168.
- TOWNSEND, J. T., & ASHBY, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge, England: Cambridge University Press.
- VICKERS, D. (1979). *Decision processes in visual perception*. New York: Academic Press.
- VICKERS, D., & PACKER, J. (1982). Effects of alternating set for speed vs. accuracy on response time, accuracy and confidence in a unidimensional discrimination task. *Acta Psychologica*, **50**, 179-197.
- WALLIS, C. P., & AUDLEY, R. J. (1964). Response instructions and the speed of relative judgments: II. Pitch discrimination. *British Journal of Psychology*, **55**, 133-142.
- WOODROW, H. (1933). Weight discrimination with a varying standard. *American Journal of Psychology*, **45**, 127-152.

NOTES

1. In the more recent North American literature, the term *semantic congruity* or *congruity* effect is often used instead of the original Audley and Wallis (1964) terminology. To this end, we henceforth use the terms *semantic congruity*, *congruity*, and *crossover* effect synonymously.
 2. The sameness-difference judgments were obtained for a different purpose and are not presented here.
 3. The response times for the pair at Level L6 provide a violation of the typical dependence of response times on both the ratio and the difference of the lengths (cf. Munsterberg, 1894; Petrusic & Jamieson, 1979). In this case, although the pairs at Levels 5 and 6 have the same ratio (32:31), the pair at Level 6 has the larger difference. Comparisons with this pair may have required additional visual scanning, thereby prolonging decision times.
 4. Complete practice-session data were available for only 7 of the 8 subjects.
 5. Discriminative accuracy, in this experiment, as in Fechner (1860/1966) and Woodrow (1933) exhibits the well-known but theoretically controversial dependence on the order of presentation. Following Jamieson and Petrusic (1975a, 1975b), we define the *presentation-order effect* by the difference in accuracy between the two presentation orders. Specifically, we subtract accuracy with the presentation order (lighter, heavier) from that with the presentation order (heavier, lighter) and then multiply this difference by one-half. The presentation-order effect is *positive* only for the very lightest of the nine different pairs, and *negative* for the remaining pairs, with the most extreme negative value obtained with the heaviest pair. Using the presentation-order effect as a measure of the *time-order error* (TOE), these data show the classic dependence of the TOE on the location of the stimulus pairs on the underlying continuum; the TOE is positive for the lightest pairs and becomes negative as the pairs become heavier. Furthermore, this dependence of the sign of the TOE on location on the attribute is invariant over the two forms of comparison used, providing a replication of the Jamieson and Petrusic (1975b) and Hellström (1977) findings with durations. This invariance over the form of the comparison shows that the TOE is not a consequence of simple peripherally based response biases (otherwise, the sign of the TOE would reverse with the form of the comparison used).
- In duration discrimination, where the properties of the TOE parallel those with weight comparisons, the prevalent theoretical position is Hellström's (1985) quantitatively developed differential sensation-weighting theory, with the weights interpretable as reflecting dynamic assimilation and contrast effects operating on the representations in memory of the stimuli to be compared. More recently, Schab and Crowder (1988) have argued that the TOE (for reproductions of paired durations) reflects the ubiquitous memory-based recency effect (but see Jamieson & Petrusic, 1975a, for evidence that more than just recency is involved). Although these various theoretical positions now make it abundantly

clear that the TOE arises as a consequence of the complex dynamic intermingling of representations in memory, it is of interest to ascertain whether these processes are precisely those involved in the semantic congruity effect. Indeed, John (1975) has suggested that the processes responsible for the TOE parallel those involved in the semantic congruity effect. However, examination of the plots in Figure 3 shows the crossover point of the semantic congruity effect to be near the 200-g weight level while the point where the TOE changes sign appears to be near the 50-g weight level. Accordingly, we conclude that the TOE and the semantic congruity effect are not mediated by identical processes. Furthermore, given the above mismatch of crossover points for the TOE and the semantic congruity effect, as well as the absence of semantic congruity effects with the proportion correct response measure in the experiment reported here, we can conclude that the data reported in John (1975, Table 1) reflect, exclusively, presentation-order effects.

6. A programming error was detected after 14 of the 20 subjects had already completed their sessions. The error caused four additional trials per block to appear with Stimulus Pair 1 (one trial with each instruction and each order of presentation). Thus, these subjects (7 in each condition) each performed a total of 400 ($100 \text{ trials} \times 4 \text{ blocks}$) experimental trials per session, while the remaining 6 subjects performed

384 ($96 \text{ trials} \times 4 \text{ blocks}$) experimental trials. Therefore, a total of 7,904 observations was obtained in this study.

7. Consider a test of the null hypothesis that semantic congruity effects do not occur against the alternative that these effects (say, for the pair with the shortest line lengths of Experiment 1 or with the most extreme of the stimuli used in Banks et al., 1982) are at least 20 msec in magnitude, which is also the value found in Experiment 4 in Petrusic and Baranski (1988a) with comparisons of supraliminally different stimuli. Suppose, conservatively, that the standard deviation of the difference in population mean response times with the two forms of the comparison is as low as 200 msec. With probability of a Type I error set at .01 (one-tailed), a difference of 20 msec would be detected 50% of the time whenever the estimate of mean response time with each instruction was based on at least 542 observations. If power is increased to a more reasonable value of, say, .90, then the minimum number of paired observations is 841. The tables provided by Howell (1982, p. 522) were used for these power calculations.

(Manuscript received June 10, 1987;
revision accepted for publication November 2, 1988.)