Journal of Experimental Psychology:
Human Perception and Performance
1976, Vol. 2, No. 3, 435-447

Semantic Congruity Effects in Comparative Judgments of Magnitudes of Digits

William P. Banks
Pomona College
Milton Fujii and Fortunee Kayra-Stuart
Claremont Graduate School

When shown a pair of digits and asked to select the larger of the two, subjects make their choice more quickly as the numerical difference between the digits increases. This article presents and tests a semantic coding model that can explain this and all previous results. In addition, this model predicts a semantic congruity effect that previous models do not predict, but which was found in both of the experiments reported here. The congruity effect occurs when subjects are asked sometimes to pick the larger and sometimes the smaller member of the pair. When the digits to be judged are both small (e.g., 2 and 3), the subjects are able to pick the smaller one faster than the larger one; when the digits are both large (e.g., 7 and 8), they pick the larger one faster than the smaller one. This model provides an alternative to image-processing models for a variety of tasks in which comparative judgments are made among elements that symbolize continuous quantities.

The amount of time required to decide which of a pair of digits is larger decreases as the numerical difference between them increases (Moyer & Landauer, 1967). This somewhat surprising phenomenon, easily replicated with adults who have had years of experience with such judgments (and who, of course, do not find them difficult at all), has been subjected to three theoretical interpretations in the past. This article advances a new model of the phenomenon and presents results that are natural predictions of the model, but cannot be handled except in an ad hoc way by the previous ones. The present and previous models are briefly sketched here, and tests of predictions of the models are presented in the final section of the article. For the sake of economy the following terms are used: *min,* to refer to the smaller member of a digit pair; *max,* to refer to the larger; and *split,* to refer to the difference between *min* and *max*. Also, digit pairs are written as two digit numbers (i.e., "12" for 1 and 2).

Two of the three previous models assume that subjects first translate each digit into an internal magnitude and then decide on the correct (larger) digit by comparing the internal representations of the two. The digit comparison is thereby converted into something like a psychophysical comparative judgment. The smaller the *split,* the smaller the difference between the internal magnitudes, and thus, the longer the time required for the decision. The first of these is the original model proposed by Moyer and Landauer (1967; see also Moyer, 1973). It assumes that the internal magnitude is an analog quantity and that it is approximately a logarithmic function of digit size.

This research was supported by a Pomona College Faculty Research Grant to the first author and by National Science Foundation Grants GU4039 and BMS 75-20328.

The authors thank the following individuals for their comments and advice on the research and on an earlier draft of this article: Dale Berger, Hiram Brownell, Paul Buckley, Herb Clark, Tim Dong, Clifford Gillman, Keith Holyoak, Marcel Just, Steve Kosslyn, Bob Moyer, Ed Smith, Lance Rips, Janet Walker, and Fred Woocher. We thank Gabriele Griesbach and Nancy Rayburn for their assistance in data collection and tabulation; the project benefited greatly from the high quality of their work.

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

The second model that postulates a translation of the digits to an internal magnitude was proposed by Buckley and Gillman (1974). Their model assumes that the internal representation of each digit is a random variable whose mean is proportional to the logarithm of the digit. The main improvement of Buckley and Gillman's model over Moyer and Landauer's is an explicit comparison process. The comparison is performed by a random walk model that begins by computing the difference between the two internal representations. If this difference exceeds a preset boundary, the indicated response is made; if the boundary is not exceeded, the process resamples the stimulus information, recomputes the difference, and adds the new difference to the old one. The process continues until the boundary is exceeded and a response can be made.

The third previous model is Parkman's (1971). This one differs greatly from the other two because it uses no nonlinear internal representations to account for the phenomena. The model assumes that when the pair of digits is presented, an internal counter begins and that the counter stops when it reaches the *min*. The subject then chooses the other digit as the larger. The model predicts a linear increase in reaction time (RT) as a function of *min*, and this prediction is upheld by Parkman's data.

Parkman's model predicts the effect of *split* on RT by reason of the fact that, with pairs of digits, *min* and *split* are confounded. The confounding is such that the sets of pairs with larger *splits* also have, on the average, smaller *mins*. The mean *min* for pairs of *split* equaling 1 (12, 23, 34, . . . , 89) is 4.5; and it is 4 for a *split* of 2, 3.5 for a *split* of 3, and so forth. If internal counting-up time is the chief determinant of RT, then RT will certainly decrease with *split*, but because of the confounded *min* variable rather than the size, per se, of the difference between the pairs.

How have these models fared against the available data? Parkman's has some problems. The basic premise that identification of digits takes place by a counting-up process predicts an increase in the simple naming latency for digits as their size increases, but such a relationship is not found (Fairbank, 1969; Theios, 1973). Also, there is still a nonlinear residual *split* effect after the effect of *min* is allowed for (Parkman, 1971), and the *min* effect itself is not always linear (Buckley & Gillman, 1974). Finally, Moyer and Landauer (1973) have shown that both Parkman's functions and their own (1967) are fitted slightly better by their equations than by Parkman's.

The models that postulate a nonlinear subjective scale of number, on the other hand, can account for a wide range of data. Probably the only available phenomena not readily explained by such models are Fairbank and Capehart's (1969) finding that choosing the larger digit is faster than choosing the smaller and Parkman's finding (1971, Figures 1 and 4) that the difference between choosing the larger and the smaller increases with *min*. These phenomena are, in fact, semantic congruity effects (discussed below), and the present model explains them satisfactorily, as will be seen.

The model proposed here has two stages of processing: (a) a first *encoding* stage, whose function it is to generate a semantic description of the stimuli—in this case, digits; and (b) a second *comparison* stage that uses the semantic codes to compute the correct response. It is assumed that the total RT for performing the task is the sum of times consumed by such stages (see Banks, Clark, & Lucy, 1975; Sternberg, 1967). Predictions for the present experiment are based only on hypothetical processes located in these two stages.

The encoding stage generates internal codes of the stimulus information presented to it, and later stages work only with such codes. In the case of comparative judgments of digit magnitudes, the size codes for the digits will of course be particularly important for processing (although other semantic information about the digits will be made available by the stage), and the primary task of this stage is assumed to be that of generating the size codes as quickly as possible. Variations in such factors as stimulus discriminability are assumed to have their effects on the time taken by the encoding stage, but it is not necessary to assume that coding

varies systematically with the size of the digit (Fairbank, 1969; Theios, 1973). It is assumed, however, that the nonlinear perception of numerical quantity (Banks & Hill, 1974; Rule, 1969; Shepard, Kilpatric, & Cunningham, 1975) causes a nonlinear mapping of digits to size codes.

The earliest size codes made available by the encoding stage are very crude: A digit is coded as either larger (L+) or smaller (S+) than a cutoff point on the numerical continuum. The cutoff is assumed to vary from trial to trial and to have a skewed distribution that places its mean below the arithmetic midpoint of the digits used in the task. The skewed distribution of cutoff placements derives from the decelerated (compressive) mapping of digits onto subjective magnitude (Banks & Hill, 1974; Rule, 1969; Shepard et al., 1975). This compression results in smaller numbers being spaced further apart on the subjective continuum than larger numbers; thus, the cutoff is more likely to fall between small than large digits.

In this model the same cutoff is always applied to both digits on a given trial. If both fall above it, the codes will be L+/L+; if both fall below it, they will be S+/S+; and if they straddle the cutoff, they will be L+/S+ or S+/L+. By assuming a common cutoff for the digits, the model never has incorrect codings emerging from the encoding stage, although they may be ambiguous, as in L+/L+ or S+/S+. The model does not, incidentally, have an explicit source of errorful responses, except for the very general one that as more processing steps are involved, there is a greater chance of an error. It thus predicts a positive relationship between RT and error rates across conditions.

The comparison stage computes the correct response by matching the previously stored instructional codes with the codes generated by the encoding stage. The instructional codes are cast in the same format as the codes for the stimuli; thus, the instruction "choose larger" is coded as L+ and "choose smaller" as S+. Because of the necessity for fast responding, the comparison stage seizes on the earliest codes available for the stimuli. If these happen to be L+/S+ or S+/L+, a match between the instructional code and one of the two stimulus codes can be made immediately, and the correct response will be quick. If, on the other hand, the stimulus codes are L+/L+ or S+/S+, the comparison stage needs more detailed codes to discriminate the digits. These more detailed codes will determine which of the large ones is larger or which of the small ones is smaller. Thus, L+/L+ is always translated to L/L+, and S+/S+ is always translated to S/S+, and then the comparison stage can select the correct response.

According to the model, the overall RT is a mix of latencies of the comparison stage resulting from S+/S+, S+/L+, and L+/L+ codings of the pair. The model predicts RT by showing how the probabilities of these codes are affected by various factors. First, the *split* effect emerges because the greater the *split,* the more likely are the digits to straddle the cutoff and be coded S+/L+ or L+/S+. The *min* effect (RT increases with *min*) comes about because the cutoff point is usually placed among the smaller numbers. Across trials, it is most likely to fall between 1 and 2, next most between 2 and 3, and so on; and it is very unlikely to fall between 7 and 8 or 8 and 9. Thus, the smaller the *min,* the greater the probability of S+/L+ coding, and the greater the probability of a response without any further recoding.

As we see, the model can predict *min* and *split* effects, but the prediction that distinguishes it from previous models is an interaction between the instructions and the overall size of the digit pair. That is, for large pairs (89, for example) the RT will be relatively faster for instructions to choose the larger one than for instructions to choose the smaller one, and the reverse will be true for smaller pairs, such as 12. This interaction is a crossover effect or, more generally, a semantic congruity effect (see Banks et al., 1975).

The semantic congruity effect emerges because the codes for the digits depend on the overall size of the pair, and processing should be faster when the codes for the digits

match the form of the instructional codes than when they do not. For example, the pair 89 is very likely to be initially coded as L+/L+ and therefore to end up coded as L/L+. If the instructions are "choose larger" (coded as L+), the correct member of the pair 89 can thus be selected faster than if the instruction were "choose smaller" (coded as S+). The S+ instruction takes longer in this case because it matches neither L nor L+, and L/L+ must be transformed to S+/S. On the other hand, a small pair (such as 12) is very likely to be coded S/S+ and will therefore lead to a faster response for S+ instructions than for L+ instructions. Note also that the *smaller* and *larger* functions will be more nearly equal for small pairs than large ones because there is more S+/L+ coding for the small pairs than large ones and, hence, less effect of congruity for the smaller pairs.

A further prediction of the model is that the size of the congruity effect will decline as *split* increases. This follows because the larger the *split*, the more likely is L+/S+ coding; and when there is L+/S+ coding, there can be no mismatch between stimulus and instructional codes and thus no congruity effect.

The present experiments seek a congruity effect in comparative judgments of digits, since the critical difference between the present and previous models is the prediction of such an effect. Both experiments therefore cross S+ and L+ instructions orthogonally with the digit pairs. The semantic congruity effect can be seen as an interaction between the instructions and the size of the pair, or more precisely, as an interaction between instructions and *min* for each *split*.

Experiment 1, in addition to testing for a semantic congruity effect, compares *splits* of 1 and 2 to determine whether the congruity effect does decline with *split* as predicted. Experiment 1 also arranges the probabilities of the various digits so that members of two pairs, 23 and 78, are unconfounded, that is, so that under both S+ and L+ instructions 2, 3, 7, and 8 are equally likely to be the chosen digit. Experiment 2 compares S+ and L+ instructions for all *splits* of 1, 2, and 3 that can be formed from the digits 1 through 9.

Method

Procedure: Experiments 1 and 2

In both experiments the subjects attempted to decide which of two simultaneously presented digits was the larger or the smaller. The digits were presented side by side in a horizontal row, and the subject had a hand-held response panel with a pair of miniature toggle switches also placed side by side. The subject indicated his choice of a digit by throwing the switch on the same side as the chosen digit.

Stimuli were presented in readouts manufactured by Industrial Electronic Engineers, in which the digits are back projected as a luminous area on a black background. The digits were approximately 1.5×2.5 cm and were placed 2.5 cm apart. The subject sat about 1 m from the readouts. A Digi-Bits logic set controlled all experimental sequences. A trial began with the experimenter saying either "choose larger" or "choose smaller" to indicate the choice the subject was to make, then throwing a switch that delivered power to the lamps in the readouts and started a timer. This procedure created a constant error in all RTs, because the rise time of the lamps was about 50 msec. The subject's response stopped the timer and extinguished the lamps, and RTs were recorded to the nearest msec. Errors by the subject caused a buzzer to sound, and stimuli that led to an error were repeated later within the experimental block in which they occurred.

Experiment 1. Experiment 1 used the digit pairs 12, 23, 13, 78, 89, and 79 equally often under both *larger* and *smaller* instructions. The pair 37 was also presented in order to equate response probabilities for 3 and 7 under the various instructions, and RTs to 37 were not recorded. If 37 had not been presented, 3 would have always been the correct digit in a pair under *larger* instructions and never under *smaller*, and the reverse would have been true for 7. The pair 37 was shown twice as often as any of the other pairs, thus making in effect eight digit pair conditions (six experimental pairs plus 37 twice). Since there were two instructions and two orders for each pair, there were a total of 32 stimulus combinations, which were given to each subject in nine separate blocks of 32. Each of the nine blocks contained a different random order of the 32 stimuli, and each subject went through the set of nine blocks in a different order. The first block encountered by each subject was considered practice, and data from it were not recorded. The reported data come from the remaining eight blocks, but their order of presentation was, unfortunately, not recorded, and examination of practice effects is not possible in Experiment 1.

Experiment 2. Experiment 2 used all pairs of the digits 1 to 9 that form *splits* of 1, 2, or 3. There were thus 21 different pairs: (a) 8 having a *split* of 1, (b) 7 having a *split* of 2, and (c) 6 having a *split* of 3. Each pair was presented equally often under *larger* and *smaller* instructions. Since each pair had two orders, there were $2 \times 2 \times 21$ or 84 conditions. These were presented in six blocks of

84, with a different random order of stimuli within each block. The first block was considered practice, and data from it were discarded. The six blocks were always presented in the same order, and effects of practice can be examined.

Subjects

Subjects were male and female students at the Claremont Colleges, paid $2 per hour. Experiment 1 had 16 subjects, of which 3 were rejected for exceeding a 5% error criterion, and Experiment 2 had 19, of which 1 was discarded for exceeding the same criterion.

Results

The effects on RT of the important variables are reported separately for each experiment, followed by sections reporting other results for both experiments at once. The standard error of the mean for some RTs is reported as a ± value in parentheses after the mean. This standard error is based on the sampling error over subjects of the mean in question. Table 1 also presents the mean RTs from both experiments.

Experiment 1

The results of Experiment 1 are shown in Figure 1, where Figure 1A plots the mean RTs for *larger* and *smaller* instructions for the pairs 12 and 89. Figure 1B plots the same functions for the pairs 23 and 78, and Figure 1C plots them for 13 and 79. Thus, Figures 1A and 1B show the results for a *split* of 1, and Figure 1C shows it for a *split* of 2. Results for the pairs 23 and 78 are plotted separately from those for 12 and 89 because 23 and 78 are the unconfounded pairs whereas the pairs 12 and 89 could possibly give results confounded by the fact that 1 is always a correct choice under *smaller* instructions and 9 is always correct under *larger* instructions.

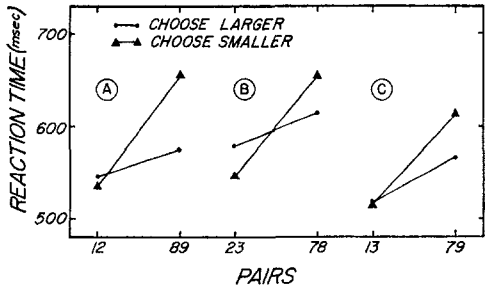

Figure 1. Reaction time to select the larger or the smaller digit of the pair indicated for Experiment 1.

TABLE 1

Mean RTs (in msec) for Each Digit Pair Under Both Instructions in Experiments 1 and 2

Experiment 1			Experiment 2		
	Instruction			Instruction	
Digit pair	Choose smaller	Choose larger	Digit pair	Choose smaller	Choose larger
12	535	548	12	559	545
23	548	582	23	588	597
13	523	522	34	611	600
78	654	616	45	591	583
89	654	578	56	665	637
79	616	566	67	635	601
			78	699	616
			89	694	590
			13	554	534
			24	545	558
			35	547	558
			46	596	576
			57	594	572
			68	609	581
			79	665	586
			14	528	551
			25	538	539
			36	569	544
			47	562	536
			58	613	549
			69	591	552

A semantic congruity effect clearly holds for each set of pairs plotted in Figure 1. In each case an interaction is observed between the size of the digit pair (12, 23, and 13 being small pairs and 78, 89, and 79 being large ones) and the instructions. This interaction is significant overall $F(1, 12) = 20.3$, $p < .01$. For the *split* of 1 alone the interaction was $F(1, 12) = 20.6$, $p < .01$, and for the *split* of 2, it was $F(1, 12) = 7.2$, $p < .025$. The size of the congruity effect decreases with *split*, as predicted by the model. The measure of the congruity effect is the mean number of milliseconds needed to move each point in order to remove the semantic congruity interaction: It is 20.4 (±4.5) msec for the *split* of 1 and 12.5 (±4.6) msec for the *split* of 2. This difference in the effect is significant, $t(12) = 3.72$, $p < .01$.

The two sets of pairs with a *split* of 1 gave very similar results. The pairs 12 and 89 were processed 22 msec faster than 23 and 78, but this difference fell short of significance, $F(1, 12) = 4.42$. The size of the congruity effect is about the same in both cases: 22.5 msec for 12 and 89, and 18.2 msec for 23 and 78 ($F = 1.03$).

The mean RT for a *split* of 1 was 589 (±23.9) msec, and it was 556 (±20.5)

msec for a *split* of 2. Testing the *split* with a planned comparison gives significance at the .01 level, $F(1, 60) = 7.36$. A test of the *split* effect was also performed in an orthogonal design, with data from the pairs 23, 78 used for the *split* of 1, omitting data from 12, 89, $F(1, 12) = 8.2$, $p < .025$. The *mins* in the experiment were 1, 2, 7, and 8; mean RTs for these were 531, 565, 613, and 616 msec, respectively. The *min* effect, by a linear contrast, is reliable, $F(1, 60) = 23.46$, $p < .01$. The best-fitting linear regression of *min* on RT ($r^2 = .84$) has a slope of 12 msec/digit, with an intercept of 528 msec.

Experiment 2

Figure 2 shows RT plotted as a function of *min* for *larger* and *smaller* instructions. Figure 2A shows this plot for a *split* of 1, Figure 2B for a *split* of 2, and Figure 2C for a *split* of 3. The semantic congruity interaction between the size of the pair and the instructions is apparent for all three *splits*. It is reliable overall, $F(20, 340) = 6.8$, $p < .01$, and is also reliable at the .01 level for each of the *splits* taken separately: $F(7, 119) = 8.4$ for a *split* of 1, $F(6, 102) = 6.4$ for a *split* of 2, and $F(5, 85) = 5.8$ for a *split* of 3.

The pairs of digits used in this experiment, as in Experiment 1, create a confounding between the effects of *min* and *split* that cannot be teased apart by orthogonal comparisons. Some of the statistical tests of *min* and *split* effects will therefore not be orthogonal. *Min* and *split* can, however, be made

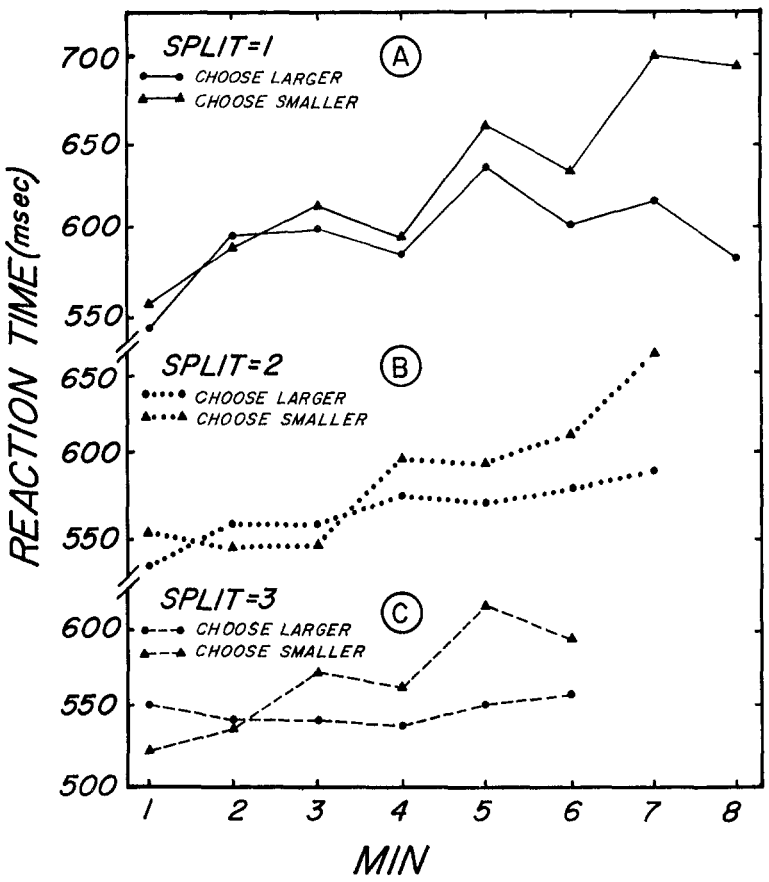

FIGURE 2. Reaction time to select the larger or the smaller digit of a digit pair, plotted as a function of the smaller member of the pair (*min*). (Data are separately plotted for the three *splits* used in Experiment 2.)

independent if all data from the pairs 78, 89, and 79 are excluded from the analysis, since dropping these pairs leaves an orthogonal 3×6 design with *splits* of 1, 2, and 3 crossed with *mins* of 1 through 6. Because this procedure loses 1/7 of the data and still leaves *max* confounded with *split,* it will be used only as an adjunct to the analysis of all the data.

In the full set of data, the mean RTs for *mins* of 1 through 8, inclusively, were 545, 561, 571, 574, 605, 595, 641, and 642 msec, respectively. The *min* effect is fairly linear ($r^2 = .60$), with a slope of 13.5 msec/digit and an intercept of 530 msec, and the linear component of the *min* effect was $F(1, 340) = 14.3$, $p < .01$. The effect of *split* is also strong with mean RTs of 613 (± 14.7), 577 (± 12.7), and 566 (± 12.4) msec for *splits* of 1, 2, and 3, respectively. The linear contrast testing the *split* effect was $F(1, 340) = 9.9$, $p < .01$. These contrasts testing *split* and *min* are, of course, not orthogonal, but there is still a *split* effect over and above the *min* effect and vice versa.

The model predicts a decrease in the semantic congruity effect as *split* increases, and the size of the congruity effect, expressed in the same terms as previously, was 14.6 (± 2.3), 9.42 (± 1.9), and 11.26 (± 2.6) msec for *splits* of 1, 2, and 3, respectively. The decrease in the effect from the *split* of 1 to 2 replicates the same comparison in Experiment 1, but the increase in the congruity effect as the *split* goes to 3 is surprising. The variation in the congruity effect with *split* in Experiment 2 is, however, not reliable, with $F < 1.0$ by a linear contrast.

The analysis of Experiment 2 that dropped the pairs 78, 89, and 79 showed significant effects of both *split* and *min,* $F(2, 45) = 45.6$ and $F(5, 85) = 24.6$, respectively. The semantic congruity effect was also reliable, $F(5, 85) = 4.8$, $p < .01$. An interaction between *min* and *split* was observed, $F(10, 170) = 5.5$, $p < .01$, in which the effect of *min* decreased as *split* increased, and the effect of *split* increased with *min.* The changes in the semantic congruity effect with *split* barely reached significance, $F(10, 170) = 2.0$, $p < .05$.

The RTs in Experiment 2 were overall a bit slower than those in Experiment 1, but the semantic congruity effect was about the same for the sets of pairs common to the two experiments. The pairs 12, 89 gave a congruity effect of 22.5 msec in Experiment 1 and 21.8 msec in Experiment 2. For the pairs 23, 78 the congruity effect was 18.2 msec in Experiment 1 and 22.5 in Experiment 2. The pairs 13, 79 gave congruity effects of 12.5 msec in both experiments.

Other Variables

As noted, information relevant to practice effects was not saved for subjects in Experiment 1, but in Experiment 2 it was, and practice did not have a reliable effect, $F(4, 68) = 2.25$, $p > .10$. The practice effect in Experiment 2 yielded mean RTs of 569, 593, 589, 584, and 598 msec for Blocks 1–5, respectively. There were no significant interactions between blocks and any other variables.

The effect of order (whether the left or right digit was the correct one) was not reliable in Experiment 1 ($F < 1.0$), but it was in Experiment 2, $F(1, 17) = 7.6$, $p < .025$, with the right side being about 20 msec faster than the left. The only reliable interaction involving order in either experiment was in Experiment 1, where order was involved in an uninterpretable three-way interaction with instructions and size of digits, $F(1, 12) = 5.7$, $p < .05$.

The instruction "choose smaller" led to RTs 20 msec slower than the instruction "choose larger" in Experiment 1 and 26 msec slower in Experiment 2. This difference is reliable both in Experiment 1, $F(1, 12) = 9.6$, and in Experiment 2, $F(1, 17) = 64.6$ (both $ps < .01$). The only reliable interactions involving instructions are the semantic congruity effects and the uninterpretable three-way interaction mentioned above.

Errors

In Experiment 1 subjects made a mean of 2.9% errors, with a range from 1% to 4.2%; in Experiment 2, 1.2% errors ranging from .1% to 2.9%. In both experiments the number of errors in each condition (summed over subjects, blocks, and order) was correlated with the correct RT: $r = .78$ in Experiment 1 and .56 in Experiment 2.

The RTs for error responses averaged 621 msec in Experiment 1 and 564 msec in Experiment 2. The error RTs were longer than the mean correct RT (578 msec) in Experiment 1, but in Experiment 2 they were shorter than the mean correct RT (585 msec). This curious reversal continues to hold up when only the correct RTs that were the make-up trials for the errorful responses are considered. In Experiment 1 the mean of these RTs was 505 msec, and in Experiment 2 it was 643 msec ($p < .01$ for the difference between error and correct RTs in both cases by binomial test). The error RTs in both experiments had zero correlations (range was $-.016$ to $+.005$) with correct RTs in those conditions as well as with the appropriate make-up RTs.

Discussion

The plots in Figures 1 and 2 confirm all but one of the qualitative predictions of the model. The semantic congruity effect is strong in each plot, and the difference between L+ and S+ instructions tends, as predicted, to increase with *min*. The size of the congruity effect decreases reliably with *split* in Experiment 1, as predicted, but in Experiment 2 it has a small variation with *split* and shows the predicted mean change only for *splits* of 1 and 2, with a reversal for the *split* of 3. This single qualitative deviation from the model is, at best, marginally reliable and, as will be seen, does not represent a large quantitative deviation from it, anyway. The other predictions of the model regarding the effects of *min* and *split* are, of course, also supported, but they are of somewhat less importance here than the congruity effects, since other models are able to predict them.

Quantitative Predictions of the Congruity Model

A least squares fit of the congruity model was made with the following equation:

$$RT = K + A[P(X/Y)_C] + B[P(X/Y)_{IC}], \quad (1)$$

where $P(X/Y)_C$ is the probability both digits will be coded congruently with the instructions (i.e., L+/L+ under "choose larger" and S+/S+ under "choose smaller"), and $P(X/Y)_{IC}$ is the probability they will be incongruent (S+/S+ under *larger* instructions and L+/L+ under *smaller*). The $P(X/Y)$'s give the proportion of the A and B latencies in the RT for a given digit pair.

The A component thus equals the time required to recode S+/S+ to S/S+ or L+/L+ to L/L+, where the instructions are congruent with the coding and no further recoding is needed. The B component should be longer than A, since it includes the same operations as the processing that requires A msec plus a congruity-creating transformation. There is also at least the logical possibility of a third latency component, C, which holds in those cases when the coding is S+/L+ or L+/S+, and no transformation of the perceptual codes is needed at all. Unfortunately, of the three possible components, A, B, and C, only two can be estimated because the associated $P(X/Y)$ values must sum to 1.0 and are therefore interdependent. Estimates of A, B, and C are thus also interdependent. Unique predictions for A, B, and C cannot be obtained in this case, but rather only an expression showing the trading relations among these factors. It was decided that, of the three components, C is most reasonably considered a part of K, the constant processing latency, and so C was omitted from the equation.

Several compressive functions were tried out for generating the $P(X/Y)$'s. Two of these, the power function with exponent less than 1.0 and the logarithmic function, were chosen because they have been proposed as the psychophysical function relating size of number to subjective magnitude (power function: e.g., Banks & Hill, 1974; log function: e.g., Rule, 1969). The exponential function was also tried, as well as a step function. These functions generated the $P(X/Y)$'s as follows. If the instructions were, for example, "choose larger," then $P(X/Y)_C$ is the probability of both X and Y being above the cutoff, and $P(X/Y)_{IC}$ is the probability of both being below it. Thus, $P(X/Y)_C$ in this case is the probability of the smaller of the two digits being coded L+, and $P(X/Y)_{IC}$ is the probability of the

larger one being coded S+. The larger or the smaller digit was selected in each case and put through the formula $P_L = f(D)$ for L+ coding and $P_S = 1 - f(D)$ for S+ coding, where $f(D)$ is one of the compressive functions, and the P_S or P_L was entered in the equation as the appropriate $P(X/Y)$. The computation of the $P(X/Y)$'s for "choose smaller" instructions was done the same way, but with $P(X/Y)_C$ being S+ coding and $P(X/Y)_{IC}$ being L+.

Log, power, and exponential functions for coding the digits fit the data almost equally well in both qualitative and quantitative terms. Table 2 gives the r^2 goodness-of-fit measure for the model with each of these coding functions. The equations shown in Table 2 contain the parameters that gave the best fit of the model in each case. Counting the three latency components (K, A, and B), the total number of free parameters required are six with a power function for the coding stage and four with either a log or exponential function.

The three coding functions resulted in predictions of *split* and congruity effects that were usually identical to the nearest millisecond. It seems, therefore, impossible to distinguish between the various psychophysical functions for digit encoding on the basis of the present experiments. The processing assumptions are apparently powerful enough to fit the data with a variety of different assumed underlying structures for the psychological number continuum (see Shepard et al., 1975, where the distinction between structure and processing assumptions is discussed).

The following predictions of effects in the experiments usually give the common prediction of the models, but are the mean of the three in the several cases where one of the predictions differed by 2 or 3 msec from the others. In each case the parenthetical quantity is the experimentally obtained value, and in only one case (Experiment 2, the congruity effect for a *split* of 2) do the predicted and obtained differ by more than 1 standard error. In Experiment 1 the functions predicted RTs of 590 (589) msec and 554 (556) msec for *splits* of 1 and 2, respectively; predicted congruity effects were 19 (20.4) msec and 17 (12.5) msec for these *splits*. For Experiment 2 predicted RTs were 604 (613), 582 (577), and 562 (566) msec, respectively, for *splits* of 1, 2, and 3; predicted congruity effects were 16.3 (14.6), 14.7 (9.42), and 13.4 (11.26) msec, respectively.

Table 2 also shows the latency components K, A, and B of the comparison stage, as expressed in Equation 1. Note that A gives the theoretical latency for a L+/L+ to L/L+ (or S+/S+ to S/S+) transformation, while B includes this time plus the time for a congruity-ensuring transformation. Thus the congruity transformation time equals the difference between A and B. The latency components with a step function for the encoding stage are not given, but the best point for the step was between 3 and 4, as expected, although the fit was still quite bad ($r^2 = .46$ and .31 for Experiments 1 and 2, respectively).

The Other Models

As mentioned in the introduction, none of the other models of processing in the task is able to account for the semantic congruity effect in digit inequality judgments. Here we consider some additional facets of the data that are troublesome for these models and try, where possible, to sketch ways in which the other models might be changed to account for the various effects they cannot otherwise account for.

Parkman's (1971) counting model might account for the congruity effect by assuming subjects sometimes count down from 10 and sometimes count up from 0. If, for example, subjects always count up when given "choose smaller" instructions and always count down under "choose larger" instructions, a semantic congruity interaction will be observed. However, such a process would have RT decreasing with *min* under *larger* instructions, in clear contradiction of the data. A more palatable assumption might be that subjects nearly always count up under *smaller* instructions but only sometimes (less than half the time) count down under *larger* instructions. A congruity effect would then be observed and RTs for both *larger* and *smaller* choices would increase with *min*. While this

TABLE 2

GOODNESS-OF-FIT FOR FOUR MODELS OF DIGIT INEQUALITY JUDGMENTS AND PARAMETERS FOR THE SEMANTIC CODING MODEL

Model	Equation[a] Encoding stage	Equation[a] Comparison stage[b]	Experiment 1[c] Comparison stage K	A	B	r^2	Experiment 2[c] Comparison stage K	A	B	r^2
Two-stage semantic coding model with digits coded according to:										
Power function	$P(\mathrm{L}+) = C(D - .9)^{.4}$	$\mathrm{RT} = K +$	344	254	312	.84	379	220	281	.67
Logarithmic function	$P(\mathrm{L}+) = C(\log D)$	$A[P(X/Y)_C] +$	392	208	260	.85	428	171	230	.67
Exponential function	$P(\mathrm{L}+) = 1 - e^{-.3D}$	$B[P(X/Y)_{IC}]$	283	306	372	.86	350	245	308	.69

Model	Equation	Experiment 1 r^2	Experiment 2 r^2
Parkman's (1971) counting model	$\mathrm{RT} = K + R(min)$	.65	.45
Counting model with separate *up* and *down* probability:			
Smaller	$\mathrm{RT} = K + R(min)$	.69	.51
Larger	$\mathrm{RT} = (K + 10pR) + (1 - 2p)R(min)$		
Moyer and Landauer (1973)	$\mathrm{RT} \propto \log(max \div split)$	.75	.66

[a] In these equations RT is reaction time, $P(\mathrm{L}+)$ is the probability of coding a digit (D) as L+, C is an arbitrary scaling constant to put the maximum $P(\mathrm{L}+)$ at 1.0, and R is the counting rate in Parkman's model.

[b] This is the same as Equation 1 in the text, with parameters defined there. Obtained values of K, A, and B for this equation are seen in the columns to the right.

[c] Semantic coding model is fit with the same encoding function but different comparison stage latency components for the two experiments; the last three models are allowed to have entirely independent parameters to fit the two experiments.

version of the model appears to give a qualitative fit to the data, it runs into serious quantitative problems, as we can see from the following considerations.

We assume that the RT under *smaller* instructions is $RT_S = K + R(min)$, where K is a constant latency and R the counting rate. The RT under *larger* instructions, with a probability p of counting down, is $RT_L = K + (1 - p)R(min) + pR(10 - min)$, assuming counting down starts at 10 (just as counting up starts at 0) and goes at the same rate as counting up. The RT_L function can be rearranged to $RT_L = (K + 10pR) + (1 - 2p)R(min)$. The predicted slope for the *larger* function is thus $(1 - 2p)R$, while it is R for the *smaller* function. Given these theoretical slopes, p can be estimated from the data as $p = (Sm - Lg)/2Sm$, where Sm is the slope of the RT function for *smaller* judgments, and Lg is the slope under *larger* judgments, and it is assumed $Sm = R$.

The first problem for this revised counting model is that p varies with *split*. It should not so vary, since the subject identifies the digits by counting and therefore cannot adjust his counting strategy on the basis of the identity of the digits. In Experiment 1, p varies only slightly, being .35 for a *split* of 1 and .26 for a *split* of 2, but in Experiment 2 p is .14, .28, and .47 for *splits* of 1, 2, and 3, respectively.

A more difficult problem for the revised counting model is the zero intercept of the *larger* function. This should be at the point $(K + 10pR)$, according to the model, but it is much smaller than this in all six functions reported here. To put the problem in more intuitive terms, the *smaller* and *larger* functions should cross according to the counting model, and the point of crossing should be at a *min* of 5. This will be the crossing point whatever the size of p, and in fact, it is the point about which the *larger* function should swivel as p varies. However, the crossing point seems to be somewhere around 2 or 3 in all the present empirical functions.

The counting model might take care of the problem with the crossover point by having a smaller value of K in the RT_L function than the RT_S function. There are two problems with this modification. First, it seems intuitively reasonable that K should be larger, not smaller, for the cases where counting down from some arbitrary point takes place. Second, K must vary with p to maintain the crossover point at about a *min* of 2, and such fiddling with K constitutes a clearly ad hoc modification of the model, since K would have to be arbitrarily adjusted to compensate for changes in p.

The counting model, in a least squares fit to the present data, gives an r^2 of .65 for Experiment 1 and .45 for Experiment 2. The fit is so poor in part because it puts a common regression line through the *larger* and *smaller* functions; it also, therefore, misses the important qualitative aspects of the data. When the probability of counting up and down is allowed to be an additional parameter (along with separate Ks for each function) in a least squares fit, r^2 increases to .69 and .51 for Experiments 1 and 2, respectively; but the number of free parameters increases from two to five, counting the point (10) at which counting down starts as a parameter. These fits are still fairly poor because they put linear functions through the evidently nonlinear *min* functions.

The random walk model of Buckley and Gillman (1974) has its greatest problem with the error latencies. A simple random walk model predicts that RTs for errors and correct responses in the same experimental condition will be equal (Pachella, 1974). Here we have unequivocal evidence that they are not, since errors are significantly longer than their correct counterparts in Experiment 1 and shorter in Experiment 2. It is odd that the errors in the two experiments should go in different directions, and perhaps the difference comes from different implicit speed–accuracy instructions. But whatever the reason for the differences between the error latencies, it is clear that a random walk model is disconfirmed by the error latencies in both experiments.

The Buckley-Gillman model would probably account for the semantic congruity effect by having different boundaries for terminal states of the random walk arrived at under larger and smaller choices, with the smaller boundary further from the starting point than the larger boundary. The com-

pressive transformation of the numbers onto the internal scale would then cause the *smaller* RT to increase more with *min* than the *larger* RT and thus to produce the effect here termed a semantic congruity effect. However, the errors come back again to plague this model. Such a difference in boundaries for the two end states would necessarily cause there to be more errors under *larger* instructions than *smaller,* since the boundary has to be closer to the starting point under *larger* than *smaller* instructions and thus will be at a lower likelihood criterion. But the errors go in the opposite direction: 3.0% and 2.1% for *smaller* and *larger,* respectively, in Experiment 1 and 1.5% and 1.0% in Experiment 2.

The quantitative fits of the Buckley-Gillman (1974) model to the present RT data depends on how the necessary additional assumptions are worked into it. Presumably, it could do a good job qualitatively, but it is difficult to know even what to expect in terms of a quantitative fit, since they do not present quantitative fits of their own data.

The original Moyer and Landauer (1967) model simply makes an analogy between digit comparisons and psychophysical comparisons of simultaneously presented physical magnitudes, and it is not really explicit enough to compare with the other models. Congruity effects are sometimes found, and sometimes not found, in psychophysical judgments (vide Banks et al., 1975), and there is no way of knowing whether the Moyer and Landauer model would predict congruity effects in the present case. However, their model does give a good fit to the data, as is seen in Table 2. This fit does not include a provision for modeling the congruity effects, but it is still fairly good because the equation does capture the nonlinear shape of the functions. However, the equation was fit separately to the two experiments. The fit of the coding model would have been somewhat better if the same freedom had been allowed it.

Some General Comments

The present results show that the internal processing in digit inequality judgments need not be done with analog quantities. That is to say, there is no need to assume that the computation of the correct response operates on an analog image, even though some of the results (*split* effect, etc.) are similar to results obtained with comparative judgments among continuous quantities in perceptual tasks. The nonlinear perception of number does play an important role in the present model, but digital codes (categorical codes as opposed to continuous, analog quantities) form the proximal stimulus to the choice stage.

The observed congruity effects make it necessary to assume the codes generated by the encoding stage have semantic properties similar to the properties of the bipolar adjectives used for the instructions. This should not be surprising, since at some point in processing the stimulus-as-coded must be in the same format as the instructions for a match to be made. In some situations (e.g., Cooper & Shepard, 1975) the instructions could be internally recoded to be stimulus-like and therefore to give evidence for analog, image-based processing. But in the present case, the instructions "choose larger" or "choose smaller" do not specify a particular digit or internal magnitude. It therefore seems by far the most efficient strategy to recode the digits to the format of the instructions (as the present model has done) rather than the reverse, or rather than recoding both to a third medium (imagery).

The various *split*-related effects, which have previously been used to argue for discriminability effects in processing (and thus to argue for an analog model), are seen here as the result of the way the coding process works. The larger the *split,* the more likely the cutoff is to fall between the digits, and therefore the more likely is a match with the instructions without further (time-consuming) recoding. In a sense, this is still a discriminability effect, but it is discriminability among the codes that determines processing, not discriminability among analog quantities. The model also accounts for the *min* effect and the *Min* × *Split* interaction by the same sort of discriminability of codes, since the decelerated function mapping digit size to coding probability places the cutoff more often among the small than the large

digits and thus allows faster processing for smaller digit pairs.

One question unresolved in this article is that of the mechanism by which the transformation of perceptual codes from L+/L+ to L/L+ and S+/S+ to S/S+ takes place. There are quite a number of ways to model this transformation, and the present experiments do not contain the sort of operations that would be needed to decide among them. It was, for example, possible to obtain very good fits to the data under the assumption that this transformation always takes a constant amount of time (about 200 msec, as it turned out) no matter what the experimental conditions. However, it seems likely that this transformation might vary with *split,* in particular, and few plausible models of the transformation process are immune to a *split* effect. But, unfortunately, it is impossible to test whether there is a *split* effect at this stage, because *split* is adequately accounted for by the model already. Introducing a *split* effect in this stage might improve the fit slightly, but the chief consequence would be simply to shift part of the explanation of the *split* effect from perceptual coding probabilities to differential durations of this recoding stage.

Finally, two points bear emphasizing: (a) The present model shows that digit inequality judgments can be included in the same theoretical context with other comparative judgment tasks, and (b) the present model gives a simple way to account for *min* and *split* effects with an approach that assumes propositional or categorical encoding. Previously, these effects have seemed to require explanation either in terms of scanning or in terms of internal comparison of self-generated perceptlike quantities (images). We feel that the present approach to modeling makes propositional encoding at least as intuitively reasonable as the other approaches to explaining these effects. Of course, the present model does make use of an analog representation, but this representation is not itself the medium in which the computation of the response is performed, as is the case with pure imagery models.

REFERENCES

Banks, W. P., Clark, H. H., & Lucy, P. D. The locus of the semantic congruity effect in comparative judgments. *Journal of Experimental Psychology: Human Perception and Performance,* 1975, *1,* 35–47.

Banks, W. P., & Hill, D. K. The apparent magnitude of number scaled by random production. *Journal of Experimental Psychology,* 1974, *102,* 353–376. (Monograph)

Buckley, P. B., & Gillman, C. B. Comparisons of digits and dot patterns. *Journal of Experimental Psychology,* 1974, *103,* 1131–1136.

Cooper, L. A., & Shepard, R. N. Mental transformations in the identification of left and right hands. *Journal of Experimental Psychology: Human Perception and Performance,* 1975, *1,* 48–56.

Fairbank, B. A., Jr. *Experiments on the temporal aspects of number perception.* Unpublished doctoral dissertation, University of Arizona, 1969.

Fairbank, B. A., Jr., & Capehart, J. Decision speed for the choosing of the larger or the smaller of two digits. *Psychonomic Science,* 1969, *14*(4), 148.

Moyer, R. S. Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics,* 1973, *13,* 180–184.

Moyer, R. S., & Landauer, T. K. The time required for judgments of numerical inequality. *Nature,* 1967, *215,* 1519–1520.

Moyer, R. S., & Landauer, T. K. Determinants of reaction time for digit inequality judgments. *Bulletin of the Psychonomic Society,* 1973, *1*(3), 167–168.

Pachella, R. G. The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition.* Potomac, Md.: Erlbaum, 1974.

Parkman, J. M. Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology,* 1971, *91,* 191–205.

Rule, S. J. Equal discriminability scale of number. *Journal of Experimental Psychology,* 1969, *79,* 35–38.

Shepard, R. N., Kilpatric, D. W., & Cunningham, J. P. The internal representation of numbers. *Cognitive Psychology,* 1975, *7,* 82–138.

Sternberg, S. Two operations in character-recognition: Some evidence from reaction-time measurements. *Perception & Psychophysics,* 1967, *2,* 45–53.

Theios, J. Reaction time measurements in the study of memory processes: Theory and data. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 7). New York: Academic Press, 1973.

(Received November 3, 1975)