Metacognition in psychophysical judgment: An unfolding view of comparative judgments of mental workload

WILLIAM M. PETRUSIC and PAULA CLOUTIER
Carleton University, Ottawa, Ontario, Canada

An experiment is reported in which it was found that when subjects were required to indicate which of two visual extents was more difficult to categorize as "long" or "short," they executed these categorizations and then measured the distance of the representation of each stimulus from the long-short category boundary; the stimulus nearer the boundary was judged to be the more difficult. When they were requested to indicate which was easier to categorize, they selected the alternative that was farther. Coombs's theory of data (1952, 1964) and his unfolding theory of preferential choice (1950, 1964) provided the conceptualization of metacognition in this psychophysical task context. Strong support for the probabilistic version of unfolding theory was obtained from the observed selective effects of laterality on the levels of stochastic transitivity attained for various classes of triples and the reliably longer times for comparisons with bilateral pairs than with unilateral pairs. The semantic congruity effects obtained, together with the changes in the form of the relationship between probability and response time as a function of practice, can be best accounted for by an evidence accrual theory in which the distances from the active reference point are measured and compared with a criterion on each evidence accrual. No support is provided for the view that propositionally based semantic "ease"-"difficulty" codes serve as the basis for these metacognitive comparative judgments of ease and difficulty.

Although theories of the process of comparing perceptual (see Link & Heath, 1975; Luce, 1986; Smith & Vickers, 1988; Townsend & Ashby, 1983), symbolic (see Banks, 1977; Birnbaum & Jou, 1990; Moyer & Dumais, 1978; Petrusic & Baranski, 1991), numerical (see Dehaene, 1989; Link, 1990), and affective (e.g., Busemeyer, Forsyth, & Nozawa, 1988; Petrusic & Jamieson, 1978) magnitudes are now well developed, the study of the more phenomenal aspects of judgmental choice behavior lags far behind. Recently, Petrusic and Jamieson (1989) initiated both theoretical and experimental investigations of the judgment of the *ease* of comparative judgments. In their experiments, on some trials, subjects were required to judge whether the stimuli in a pair of brightness patches were the same or different, and on other trials, subjects directly compared two pairs of brightness patches, select-

ing the pair in which it was easier to determine sameness-difference.

Each of Petrusic and Jamieson's (1989) 4 observers was able to make these quaternary relational judgments (comparisons of pairs of pairs) in an orderly and consistent manner. Indeed, Petrusic and Jamieson showed that these judgments satisfied the axioms for a positive difference structure (Krantz, Luce, Suppes, & Tversky, 1971).¹ Consequently, the quaternary relational judgments of ease permitted representation of the stimuli such that, whenever the judgment with the pair (a, b) was easier than that with the pair (c, d) (denoted $abEcd$), the difference in the representations of a and b , $s(a) - s(b)$, was at least as large as the difference between the representations of stimuli c and d ; that is, $s(a) - s(b) \geq s(c) - s(d)$ if and only if $abEcd$.

Despite these encouraging findings, it remains unclear whether the quaternary comparisons of the ease of comparative judgment were based exclusively on the phenomenal ease of judgment. They may well have been based strictly on differences in sensory magnitudes, rather than on the more phenomenal aspects of the comparison process. In particular, the quaternary comparisons of ease may not have included the requisite sameness-difference comparison, and they may have been based on a similarity-dissimilarity comparison.

In the present report, we provide further explorations of the metacognitive aspects of psychophysical judgment. In particular, we required subjects, on some trials, to cate-

This research was supported by a Natural Sciences and Engineering Research Council grant to W. M. Petrusic. The data reported here were collected as part of an honors thesis submitted by Paula Cloutier to the Psychology Department at Carleton University. Completion of the article was made possible through the marvelous hospitality of the Department of Psychology at the University of Western Australia during Petrusic's study leave. Christine Gatti, Geoffrey Hammond, John Hogen, Chris Pratt, and Tony Gibbs provided hardware, software, and general support throughout. We also thank Joseph Baranski, Jo-Anne LeFevre, Neil MacMillan, and an anonymous reviewer for numerous informative and critical comments. Address reprint requests to William M. Petrusic, Department of Psychology, Carleton University, Ottawa, ON, K1S 5B6, Canada.

gorize a single visual extent as "long" or "short," and on other trials, to judge which alternative in a pair of simultaneously presented visual extents was *easier* to categorize as long or short. In addition, we used both forms of the comparison, and on one half of the comparison trials, we required subjects to select the line segment that was *more difficult* to categorize as long or short. We thought that this task would overcome the difficulties mentioned above with the Petrusic and Jamieson (1989) task and would thus permit direct and detailed examination of the decision processing and representation of *self-generated* information.

This experiment was also motivated by Coombs's (1950, 1964) pioneering abstract view of how observations arising from very different tasks and in diverse problem domains can be mapped into the same formal relational system and then analyzed to obtain a deep understanding of the cognitive processes and structures underlying the observations. In the present instance, Coombs's *unfolding theory of preferential choice* behavior provides the underlying formal representation of the choice processes involved in binary comparisons of the ease-difficulty of binary categorization. Consequently, elegant and very powerful tests become available to assess, directly, the representability of these more phenomenal aspects of judgmental choice behavior—notably, the ease and the difficulty of binary categorizations—in terms of Coombs's formal data-measurement system.

According to Coombs's (1950, 1952, 1964) unfolding theory of preferential choice behavior (see also Coombs & Avrunin, 1988), subjects' ideal points and stimuli can be simultaneously represented on a continuum referred to as a Joint or J-Scale. Preference orderings, or I-Scales, as they are called, are generated by folding the J-Scale at the subject's ideal point, or point of maximum utility. That is, for any pair of stimuli, x and y , x is chosen over

y if and only if x is closer on the J-Scale to the subject's ideal point, I . Figure 1 provides an example of a J-Scale folded at the ideal point, I , along with the resultant preference ordering, or I-Scale.

In the present context, the ideal point represents the stimulus that is the most difficult to categorize as "long" or "short" in the continuum—namely, the category boundary between long and short. Selection of the stimulus that is *more difficult* requires the subject to compare the distance of each stimulus from the category boundary and to select the stimulus that is *nearer* to the category boundary (the ideal point). On the other hand, selection of the stimulus that is *easier* to categorize requires the subject, in unfolding-theory terms, to choose the alternative that is *farther* from the ideal point. Thus, on this view, comparative judgments of both the ease and the difficulty of the binary categorizations must be subject to the internal constraints on relations among choice probabilities and response times, which are imposed formally by the structural relations inherent in Coombs's probabilistic version of his unfolding theory. These constraints will be especially evident in the levels of stochastic transitivity satisfied by the various classes of triads distinguished initially by Coombs (1958).

Strong stochastic transitivity (SST) is satisfied as follows: If for all triples a , b , and c in the choice set A , $P(aRb) \geq .5$ and $P(bRc) \geq .5$, then $P(aRc) \geq \max[P(aRb), P(bRc)]$, where $P(aRb)$ is the observed probability that stimulus a is in relation R to stimulus b ; that is, a is nearer the ideal point than stimulus b is, and in the present context, a is harder to categorize than stimulus b is. Thus, in this condition, the choice probability with the bounding pair in the triad cannot be smaller than either of the probabilities for the embedded pairs. A weaker condition, *moderate stochastic transitivity* (MST), is defined as follows: if $P(aRb) \geq .5$ and $P(bRc) \geq .5$, then

Figure 1. The underlying unidimensional attribute or Joint Scale (J-Scale), which includes the subject's ideal point, I , and the resultant preference ordering (I-Scale) of proximities of the stimuli to the ideal point. The I-Scale is obtained by folding the J-Scale at the ideal point, I .

$P(aRc) \geq \min[P(aRb), P(bRc)]$. The weakest condition, merely asserting the existence of an ordering, referred to as *weak stochastic transitivity* (WST), is given thus: if $P(aRb) \geq .5$ and $P(bRc) \geq .5$, then $P(aRc) \geq .5$.

Various classes of triples arise on the basis of the laterality relations of the component pairs. The pairs are termed *bilateral* when they are on the opposite sides of the ideal point (e.g., stimuli *a* and *g* on the J-Scale shown in Figure 1) and *unilateral* when they are on the same side of the ideal point (e.g., *a* and *b* on the J-Scale shown in Figure 1). When all the stimuli in the triple are on the same side of the ideal point, the triple is called *unilateral* (e.g., the triples *def* and *cba* shown on the I-Scale given in Figure 1). Coombs (1958) distinguished two classes of bilateral triples: *bilateral split* (e.g., the triple *ceb* in Figure 1), where the single alternative, stimulus *e*, on the opposite side of the ideal point is between (splits) the other two alternatives (the unilateral pair *c* and *b*) in the preference ordering, and *bilateral adjacent*, where the single alternative on the opposite side of the ideal point is either last (e.g., the triple *deb*) or first (e.g., the triple *dcb*) in the preference ordering. Bechtel (1968) further distinguishes the bilateral adjacent triples as *bilateral above* (e.g., the triple *deb*) or *bilateral below* (e.g., the triple *dcb*), depending on whether the single alternative is last or first in the preference ordering (whether the single opposite-signed stimulus folds, respectively, above or below, the two like-signed stimuli in the terminology developed by Bechtel, 1968).

According to the probabilistic version of Coombs's unfolding theory (Coombs, 1958, 1964; Coombs, Greenberg, & Zinnes, 1959), trial-to-trial variability in the location of the subject's ideal point, *I*, will exert selective effects on the consistency of binary choices, depending on whether the pair is bilateral or unilateral. Since variability in the ideal point changes the relative proximity of the stimuli to the ideal point in a bilateral pair, ideal point variability adds to the variability in the stimulus representation of bilateral pairs. However, variability in the ideal point adds considerably less variability to the stimulus representation for unilateral pairs than it does to that for bilateral pairs. Consequently, consistency of choice will generally be higher for unilateral than for bilateral pairs.

These selective effects of ideal point variability on bilateral and unilateral pairs will naturally exert profound effects on the transitivity relations that must hold among the three classes of triples defined above according to Coombs's probabilistic unfolding theory. The bilateral-split triples should rarely violate SST, since any effect of ideal point variability will be greater for the embedded bilateral pairs (*dc* and *ce*) than it will for the bounding unilateral pair (*de*). On the other hand, the bilateral-adjacent triples should rarely satisfy SST; in the triple *deb*, for example, ideal point variability will have a greater effect on the bounding bilateral pair (*db*) than it will on the embedded unilateral pair (*de*). Finally, the unilateral triples should satisfy SST at a level intermediate to that for the bilateral-split and the bilateral-adjacent pairs.

Bechtel's (1968) version of unfolding theory, which is especially useful for obtaining estimates of the locations of the stimuli and the ideal point on the J-Scale with group data, permits variability only in the ideal point. The transitivity relations among the classes of triples predicted by this theory are as follows: unilateral triples satisfy exactly MST, bilateral splits satisfy at least MST, bilateral-below triples satisfy exactly MST, and bilateral-above triples satisfy at least MST. Bechtel also proved that bilateral split triples would satisfy SST whenever the distribution of ideals is symmetric.

To extend this investigation of metacognition in psychophysical judgment further, the present experiments also provide a series of analyses of response times as an aid to our understanding of the decision rules underlying these previously unexplored phenomenal ease comparisons. The examination of response times in these binary ease-difficulty comparisons takes two different but interrelated directions. First, since the present ease-difficulty comparisons are interpretable as preferential choice data, the response-time-response-probability data permit examination of some current theoretical positions and methodological issues in the study of probabilistic preferential choice behavior. Second, since we require subjects to select, on some trials, the alternative that is easier to categorize and, on other trials, the alternative that is more difficult to categorize, the present experiment permits a study of *semantic congruity effects* with these preferential-choice-phenomenal-ease comparisons.

Congruity effects reflect an interaction between the form of the comparison and the location of the alternatives to be compared on the relevant continuum. In the present context, congruity effects would be evident if selection of the alternative "easier to categorize" were faster than selection of the alternative that was "harder to categorize" when the alternatives in the pair were "easy" to categorize. Conversely, selection of the alternative that was "harder" would be faster when the alternatives were both difficult to categorize. Although congruity effects were first evident in the experiments of Shipley, Coffin, and Hadsell (1945) and Shipley, Norris, and Roberts (1946) concerning the pleasantness and unpleasantness of colors, respectively, they remain to be systematically explored in a preference task.²

The relationship between the probability of a particular response and the average time for that response to occur, referred to as the *latency probability function* (LPF), provides a first-order test of the various quantitative theories admitting response-time-response-probability relations (see, e.g., Audley & Pike, 1965; Luce, 1986; Morgan & Robertson, 1980; Petrusic, in press; Petrusic & Jamieson, 1978; Pike, 1968; Vickers, 1979; Vickers, Caudrey, & Willson, 1971). Petrusic and Jamieson (1978), in their examination of replicated paired comparison preferential choice, showed that the form of the LPF is dependent on both the effects of practice and the importance of the decision. In order to examine the effects of practice on the form of the LPF, they obtained latency

frequency curves (LFCs), empirical LPFs with choice instance held constant. LFCs were strictly monotonically decreasing on the *first choice instance* for each of the three choice sets they examined: a priori important decisions involving two-outcome lotteries, statements of opinion concerning fraternities, and a priori unimportant decisions concerning the aesthetic appeal of isosceles triangles. The LFC remained monotonically decreasing (even though response times decreased) only for the set of a priori important decisions. If, on the other hand, the decisions required were a priori unimportant the LFC became non-monotonic for all subsequent choice instances following the first. Petrusic and Jamieson (1978) accounted for this particular dependence of the form of the empirical LPF on the importance of the decision and stage of practice in terms of a discrete state evidence accrual model that permits both slow and fast guessing. When decisions are important, the accrual criterion for guessing is high and the LPF is monotonically decreasing. On the other hand, as responses increase in speed, with the unimportant decisions, the accrual criterion for fast guessing is lowered and the model predicts the observed nonmonotonic LPF.

Recently, Busemeyer et al. (1988) derived response time properties of Restle's (1961) *suppression of aspects* (SOA) and Tversky's (1972) *elimination by aspects* (EBA) models. They have shown that the LPF is strictly monotonically decreasing for the SOA model but flat for the EBA model (i.e., response times are constant and are independent of the probability of the occurrence of the response). The present experiment permitted a test among these three classes of theories. In it, we obtained a large number of replications with each stimulus pair over a number of experimental sessions. Consequently, we were able to obtain LPFs, separately, for each successive block of trials, in order to provide a test of applicability of the SOA, the EBA, and the slow and fast guessing models in the context of these phenomenal ease-difficulty comparisons. If the phenomenal representations of ease-difficulty are in fact multifaceted, of course the SOA and EBA theories can in principle provide viable accounts of the present data. However, if, as we have argued, the judgments are based on Coombs's unidimensional theory of preferential choice, the LPF might well be strictly monotonically decreasing during the initial block of choice instances but not necessarily monotonically decreasing on subsequent blocks.

METHOD

Subjects

Four undergraduate students (2 males and 2 females) were paid an hourly rate to serve for 11 experimental sessions. All were naive concerning psychophysical experimentation.

Stimuli and Design

Nine horizontal line segments (12.48, 32.45, 52.42, 72.39, 92.36, 112.33, 132.30, 152.27, and 172.24 mm; labeled 1 through 9, respectively) were used as stimuli in the two judgmental tasks. They were presented on a video monitor.

The *binary categorization* task required the subjects to classify each of the nine line segments as either long or short. Twenty-five blocks with each of the nine stimuli occurring six times in each block of 54 trials were obtained over the course of the first five experimental sessions. Thus, overall, each of the nine stimuli was presented for categorization for a total of 150 replications. A different random order of presentation was used for every block of 54 trials.

The *comparative judgment* task required the subjects to indicate which of two simultaneously presented horizontal line segments (centered on the screen both horizontally and vertically) was easier to categorize as long or short on one half of the trials; on the other half of the trials, the subjects selected the line segment that was harder to categorize. The stimulus pairs consisted of the 36 pairwise combinations of the nine line segments. Over each block of 144 trials, each of the 36 pairs appeared once with each instruction, with the position of each element in the pair balanced over the top and bottom of the screen. Binary comparisons were obtained over the course of six experimental sessions, with each session comprising four blocks of 144 trials. A different random ordering of the 144 trials was used in each block. Consequently, over the six sessions, 48 replications were obtained with each pair with each instruction, after collapsing over the nuisance positional variable.

Apparatus

The stimuli and instructions were presented on an Amdex-310A video monitor under the control of an IBM-XT-compatible computer. All other events were also under computer control, including sequencing, timing, and recording of responses within 1-msec accuracy. Situated directly to the right of the monitor was a response panel (a PC mouse) with three response buttons. A label indicating the function of the buttons was affixed to the mouse and varied for each of the two tasks. In the binary categorization task, the top button on the response panel was labeled "long" and the bottom button was labeled "short." In the comparative judgment task, the top and bottom buttons of the response panel corresponded to the location of the two simultaneously presented horizontal line segments.

Procedure

Detailed instructions, explaining the nature of the display and the task at hand, were given to each subject. All subjects were instructed to respond as quickly as possible without being careless. Both speed and accuracy were stressed. The experimental sessions took place in a small room in which the subject was seated at a table that supported the computer terminal and the response panel. The terminal, located 50 cm in front of the subject, displayed the stimuli at eye level.

In the binary categorization task, each trial was initiated with the presentation of the instruction "line length categorization" 500 msec prior to presentation of the line length. On one half of the comparative judgment trials, the instruction "which is harder to categorize?" appeared on the screen for 500 msec prior to the presentation of the stimulus pair, and the instruction "which is easier to categorize?" preceded stimulus pair presentation on the remaining half of the trials. In both cases, the instruction for the comparative judgment, along with the stimulus pair, remained on the screen until the subject responded.

RESULTS AND DISCUSSION

We will present the findings in six sections. The first, preliminary to examination of the comparative judgments, provides a view of performance with the binary categorization task. In the second, we consider whether these phenomenal comparisons can indeed be viewed as uni-

dimensional unfolded preferential choices; this section includes an examination of the potential selective effects of laterality on stochastic transitivity. The third provides an exploration of the response time properties of the comparative judgments, and in the fourth, the response time analyses are extended to an examination of semantic congruity effects. The fifth is an examination of the relations between response times in the categorization and the comparison tasks. We conclude with an examination of the form of the LPF as a function of practice.

Binary Categorization

Figure 2 provides, separately for each subject, plots of the probability of the response "long" as a function of the length of the stimulus and, at the same time, overall median response times with each stimulus. For each subject, these psychometric functions are well behaved, monotonically increasing with the extent of the (variable) stimulus. The "50% point," which provides an estimate of the location of the category boundary between "long" and "short," is given by some intermediate point along the extent continuum for each of the 4 subjects.

The plots also show that response time increases as the uncertainty of the categorization increases; response times are maximal at the stimulus categorized as "long" approximately 50% of the time, precisely as is evident in

the landmark binary categorization data of Cartwright (1941). Taken together, the plots in Figure 2 show that these subjects performed the binary categorization task as required.

Phenomenal Comparisons of the Ease and the Difficulty of Binary Categorization as Unfolded Preferential Choice

For each subject, we obtained an ordering of the difficulty of categorization of the stimuli in the set on the basis of the *stochastic dominance* (SD) criterion. Stimulus *a* is stochastically dominant over *b* whenever it is selected more than half of the time. If we let $P(aEb)$ denote the probability that *a* is judged to be easier to categorize than *b* and $P(aDb)$ denote the probability that *a* is judged to be more difficult than *b* to categorize, then $aSDb$ whenever $\{P(aDb) + [1 - P(aEb)]/2\} > 0.50$. Estimates of $P(aEb)$ and $P(aDb)$, obtained from six sessions of 576 trials in each session, were based on 48 replications, and overall, after the data obtained with the two instructions were combined, the stochastic dominance criterion was based on 96 observations for each pair.³

As stated earlier, weak stochastic transitivity (WST) provides the defining condition for the existence of an ordering of the stimuli in the choice set with replicated choices. In terms of the stochastic dominance criterion,

Figure 2. Psychometric function and response times for the binary categorization task; filled circles provide the probability of the categorization "long" (right ordinate). Unfilled circles indicate the median overall response time with each stimulus for the categorization task (left ordinate). The upper plots provide overall response times for the comparison task with the instruction to select the stimulus that is "easier" to categorize (filled triangles) and "harder" to categorize (unfilled triangles) for comparisons with the stimulus nearest to the point of subjective equality.

Table 1
Obtained Stochastically Dominant Orderings (I-Scales), From Most to Least Difficult to Categorize as "Long" or "Short"

Subject	Obtained Ordering
D.L.G.	8 7 9 6 5 4 3 2 1
M.A.J.	6 5 7 4 3 8 2 9 1
N.L.B.	5 6 7 8 4 9 3 2 1
P.A.D.	6 7 5 8 9 4 3 2 1

Note—The stimuli are labeled 1–9, from shortest to longest.

WST can be expressed as follows: if $aSDB$ and $bSDc$, then $aSDc$. Importantly, WST was satisfied for each subject in every one of the 84 triples providing tests. Thus, there can be no question that each of these subjects was able to order the stimuli in terms of the ease and the difficulty of binary categorization.

Furthermore, examination of the rank orderings of difficulty (based on the stochastic dominance criterion) provided in Table 1 also shows that each ordering can be viewed as arising from folding the a priori J-Scale at the ideal point, which in this case corresponds to the hypothetical category boundary between long and short lines.⁴

It is also of interest to note that for each subject, the "50% point" on the categorization psychometric function is in fact between the two stimuli that were judged to be the most difficult to categorize. For Subject M.A.J., for example, the "50% point" is between Stimuli 5 and 6 (Figure 2), and those two stimuli are the most difficult to categorize (Table 1). Furthermore, the rank ordering of the response times in the categorization task (see Figure 2) permits accurate prediction of the ordering of difficulty obtained with the comparison task. For Subjects M.A.J., N.L.B., P.A.D., and D.L.G., the comparison-

based ordering was correctly predicted in 32, 34, 35, and 31 of the 36 pairs, respectively, with the rank ordering of the categorization times.

The Selective Effects of Ideal Point Variability on Stochastic Transitivity

The classification of triples for the analysis of transitivity was based entirely on the rank ordering of difficulty according to the stochastic dominance criterion. The stimulus nearest to the ideal point (the category boundary) was excluded because its location relative to the boundary could not be determined with any degree of certainty (see also Coombs, 1958, p. 5). Consequently, the laterality relations with the set of 28 pairs formed from the remaining eight stimuli provided the basis for the classification of the resulting 54 triples. Table 2 summarizes the examination of stochastic transitivity for each of the 4 subjects.

The selective effects of laterality on stochastic transitivity are clearly evident from these phenomenal ease and difficulty comparisons for each of the 4 subjects. For subject M.A.J., for example, SST is violated but once with the 15 bilateral-split triples. On the other hand, only 5 of the 25 bilateral-adjacent triples satisfy SST. Over all subjects, SST is satisfied in 40 of the 42 bilateral-split triples but it holds for only 33 of the 117 bilateral adjacent triples [$\chi^2(1, N = 159) = 59.83, p < .0001$]. Thus, largely as predicted and obtained by Coombs (1958), SST is rarely violated when triples are bounded by a unilateral pair, especially if a bilateral pair is embedded in the triple. On the other hand, SST is rarely satisfied for the bilateral-adjacent triples, because the embedded unilateral pairs in each triple are bounded by a bilateral pair. These selective effects of laterality on SST provide strong sup-

Table 2
Levels of Stochastic and Temporal Transitivity Satisfied by Each Class of Triple

Subject	Class of Triple	<i>n</i>	SST	MST	STT	MTT
M.A.J.	bilateral-split	20	0.95	1.00	0.95	1.00
	unilateral	11	0.45	0.73	0.64	1.00
	bilateral-above	13	0.15	1.00	0.54	0.92
	bilateral-below	12	0.25	1.00	0.25	0.83
P.A.D.	bilateral-split	10	1.00	1.00	1.00	1.00
	unilateral	11	0.45	0.73	0.64	0.91
	bilateral-above	12	0.75	1.00	0.50	1.00
	bilateral-below	23	0.13	1.00	0.09	0.83
D.L.G.	bilateral-split	6	0.83	1.00	1.00	1.00
	unilateral	35	0.40	0.83	0.51	0.91
	bilateral-above	—	—	—	—	—
	bilateral-below	15	0.33	1.00	0.06	0.67
N.L.B.	bilateral-split	6	1.00	1.00	1.00	1.00
	unilateral	8	0.38	0.75	0.50	0.88
	bilateral-above	21	0.29	0.95	0.57	1.00
	bilateral-below	21	0.38	1.00	0.29	1.00
Over all subjects	bilateral-split	42	0.95	1.00	0.99	1.00
	unilateral	65	0.42	0.76	0.57	0.93
	bilateral-above	46	0.30	0.74	0.42	0.73
	bilateral-below	71	0.27	1.00	0.17	0.83

Note—Cell entries are proportions of triples in each class satisfying the specified level of transitivity. SST and MST denote strong and moderate stochastic transitivity, respectively, and STT and MTT denote their respective temporal equivalents.

port for the view that the phenomenal ease and difficulty comparisons can be conceptualized in terms of Coombs's (1950, 1952, 1964) unfolding theory of preferential choice behavior. Moreover, as noted above, Bechtel's (1968) probabilistic unfolding theory requires that unilateral triples must satisfy exactly MST. However, the fact that a substantial number of the unilateral triples satisfy SST suggests that Bechtel's version of unfolding theory is not appropriate for the present individual subject data.⁵ Rather, this model appears to be more appropriately suited for group data (e.g., Petrusic, Cousins, & Corbin, 1984) than for individual subject data. In summary, it is clear that the present findings are in accord with earlier demonstrations of laterality by Hall (1971), Hall and Weir (1974), and Petrusic and Jamieson (1976) and with the conclusions that they reached.

Response Time Analyses of Ease and Difficulty Comparisons

Distance effects with a fixed standard. The upper plots in Figure 2 provide a view of the properties of response times from the paired comparison task. Median overall response times are given, separately for each instruction, for each of the comparisons with the stimulus nearest to the category boundary viewed as the standard according to the method of constant stimuli. As in the classic Johnson (1939) experiment, for example, response times are generally decreasing, exhibiting the typical robust dependence of comparison time on the differences in perceived magnitude when the method of constant stimuli is employed.

Laterality and distance effects. The effects of laterality are clearly evident in the distance effect plots provided in Figure 3. For each subject, comparisons with both the unilateral and the bilateral pairs exhibit the typical dependence of response times on the number of ordinal steps between the elements in the pair—that is, there was a distance effect. In addition, the plots in Figure 3 show that, for each subject, response times were uniformly longer with the bilateral pairs than with the unilateral pairs in every instance with approximately comparable distances.

Using the median response time with each pair at each level of ordinal distance and with each instruction, for both the unilateral and the bilateral pairs, as the dependent variable, analyses of variance were conducted on the data for each subject. These analyses of variance showed that the main effect of distance was statistically reliable for each of the 4 subjects [$F(3,40) = 10.87, p < .0001$; $F(2,38) = 30.53, p < .0001$; $F(3,38) = 4.61, p < .008$; and $F(5,32) = 7.44, p < .0001$, for Subjects M.A.J., N.L.B., P.A.D., and D.L.G., respectively]. In addition, bilateral comparisons were reliably longer than unilateral comparisons for each of the 4 subjects [$F(1,40) = 25.05, p < .0001$; $F(1,38) = 45.04, p < .0001$; $F(1,38) = 33.25, p < .0001$; and $F(1,32) = 41.62, p < .0001$, for Subjects M.A.J., N.L.B., P.A.D., and D.L.G., respectively].⁶

Laterality effects and temporal transitivity. Petrusic and Jamieson (1976), in their examination of response-time-response-probability relations in preferential choice

Figure 3. Mean of median overall comparison times, as a function of the ordinal separation between the elements in the pair, compared for each subject for bilateral (circles) and unilateral (triangles) pairs.

behavior, established that the selective effects of laterality on stochastic transitivity were also evident with response times. To perform these response-time-based analyses of probabilistic transitivity, we defined the temporal equivalent of SST, *strong temporal transitivity* (STT), as follows: for all alternatives a , b , and c in the choice set, if $aSDb$, $bSDc$, and $aSDc$, then $RT(a,c) \leq \min[RT(a,b), RT(b,c)]$. Thus, for each triple ordered on the basis of SD, response times with the bounding pair will be faster than response times with either of the two embedded pairs in the triple. The temporal equivalent of MST, *moderate temporal transitivity* (MTT), requires that the response time with the bounding pair not be the slowest; that is, $RT(a,c) \leq \max[RT(a,b), RT(b,c)]$. The temporal transitivity analyses, also provided in Table 1, are based on the median of the overall response times. Over all subjects, there are 42 bilateral-split triples and STT is satisfied in 41 of these. On the other hand, STT is satisfied with only 31 of the 117 (26.4%) bilateral-adjacent triples [$\chi^2(1, N = 159) = 63.14, p < .0001$]. Thus, the selective effects of laterality on STT parallel, precisely, those obtained with the more typical SST condition. Taken together, these analyses provide strong evidence that the phenomenal comparisons of ease and difficulty of categorization arise from folding the extent dimension at the category boundary, that both stimulus and ideal point variability contribute to comparisons with bilateral pairs, and that only stimulus variability is involved with unilateral pairs.

Semantic Congruity Effects

Figure 4 provides plots of response times averaged over all 4 subjects (means of individual subject medians), for each of the adjacent (one-step) pairs with each instruction. These plots show that these ease and difficulty comparisons were subject to the semantic congruity effect. When subjects were instructed to select the alternative that was harder to categorize, comparison was faster for alternatives that were in fact difficult to categorize. Conversely, with pairs that were relatively easy to categorize, selection of the alternative that was easier to categorize was faster than selection of the alternative that was more difficult to categorize.

Using the median response time over the eight replicates in each of the six sessions with each of the pairs adjacent in the difficulty ordering and each instruction as the dependent variable, an analysis of variance with this $8 (\text{pairs}) \times 2 (\text{instructions}) \times 6 (\text{sessions})$ factorial, entirely within-subject design was conducted. In this analysis, the orthogonal polynomials were modified to accommodate the unequal spacing of the pairs (midpoints) along the length continuum, and trend analyses were conducted. A significant linear trend [$F(1,3) = 10.80, p < .046$] confirmed the increase in comparative judgment times with increases in the underlying difficulty of the component categorizations. The interaction between pair location and instruction, attesting to the overall reliability of the semantic congruity effect attained significance by con-

Figure 4. Mean of median response times over all subjects with each instruction for the comparison task for pairs adjacent in the ordering of difficulty (from easy to hard, based on the stochastic dominance criterion). Averages over subjects were obtained by averaging over adjacent pairs in the rank ordering of each subject. Filled triangles are for the instruction to select the element that is "harder" to categorize, and unfilled triangles are for the instruction to select the element that is "easier" to categorize. Filled circles provide average times for the categorization task.

ventional test [$F(7,21) = 2.68, p < .038$] but not with Greenhouse-Geisser epsilon-adjusted degrees of freedom.⁷ Finally, neither the main effect of instruction ($F < 1.0$) nor that of sessions ($p > .08$) attained statistical significance.

The individual subject plots provided in Figure 5 show that the preceding conclusions are not an artifact of averaging response times over subjects. Subjects M.A.J., P.A.D., and D.L.G. exhibit the full "crossover" effect, and Subject N.L.B. shows a limited form of the effect referred to as the "funnel" effect (see Audley & Wallis, 1964). Separate analyses of variance were also performed on the individual subject data. These analyses parallel those reported above for the group data, except that replicates were obtained by using median times with each pair and each instruction for each of the six sessions.

Although congruity effects are clearly evident in Figure 5 for Subject D.L.G., the high variability in this subject's response times precluded establishing their statistical reliability [$F(1,5) = 3.34, p > .127$, for the interaction of instruction with the linear trend]. However, statistically reliable semantic congruity effects were established for each of the 3 remaining subjects. For Subject M.A.J., the interaction of the linear trend component with the pair location factor and instruction was highly reliable [$F(1,5) = 42.84, p < .001$]. For Subject P.A.D., reliable congruity effects were evident with the significance of the interaction with instruction of the quartic trend over pair location [$F(1,5) = 6.83, p < .047$]. Interestingly, for Subject N.L.B., instruction interacted reliably with the quadratic trend component [$F(1,5) = 8.30, p < .034$], confirming the enhancement of the funnel effect for the bilateral pairs, as shown in Figure 5. It is also of interest

Figure 5. Median of overall response times with each instruction for pairs adjacent in the ordering of difficulty (from easy to hard, based on stochastic dominance) for each subject. Unfilled points are for selection of the element in the pair that is "harder" to categorize, and filled points are for the instruction "easier." The plot also indicates whether a pair was unilateral (triangles) or bilateral (circles).

that the effect is especially clear for Subject M.A.J.; each of the pairs adjacent in the difficulty ordering involved bilateral comparisons. Thus, these plots and analyses establish and extend the influence of decision difficulty on the magnitude of the congruity effect to include the effects of laterality; congruity effect magnitude is enhanced when the comparisons involve bilateral pairs.

Stochastic dominance and congruity magnitude. Petrusic (in press) and Petrusic and Baranski (1989a, 1989b) found larger congruity effects with error responses than with correct responses whenever error response times were longer than correct response times. In the present context of phenomenal ease and difficulty comparative choice data (as in binary preferential choice tasks), of course, responses are neither correct nor incorrect. Rather, generally, one of the two responses with each pair (and each instruction) occurs more frequently than the other, and, as indicated earlier, this response is defined as the *stochastically dominant* response. It is natural to determine whether differential congruity effects, paralleling those obtained in tasks with correct and incorrect responses, are obtained with the stochastic dominance criterion. The question is important, because in addition to extending the generality of the properties of congruity effects, alternative theories of the semantic congruity effect are distinguished on the basis of the obtained relation.

Figure 6 provides plots of the semantic congruity index, separately for the set of stochastically dominant and

the set of stochastically nondominant responses. The index, obtained by subtracting average (mean of medians) response times with the instruction "easier" from response times with the instruction "harder," is positive for the "easy" pairs and negative for the "hard" pairs for *both* the stochastically dominant and the stochastically nondominant sets of responses. These indices, in addition to exhibiting substantial congruity effects with both classes of responses, also show that congruity effects are uniformly larger with the set of stochastically nondominant responses ($p < .05$, by sign test).

Response probability and the congruity effect. The entries in Table 3 provide response probabilities for each pair, adjacent in the ordering of difficulty, with each instruction. Examination of the entries in Table 3 shows that congruity effects are not clearly evident for any one of the 4 subjects. It is not the case that the probability of response was greater with the instruction "harder" than with the instruction "easier" when the pairs were judged to be relatively difficult. Similarly, response probabilities were not higher with the instruction "easier" for the easy pairs (i.e., those ranked last in terms of difficulty). Unfortunately, clear tests are not possible because of the extremely high level of consistency in responding with both instructions, especially with the relatively easy pairs. Subject N.L.B. appears to exhibit a congruity effect, and for Subjects D.L.G. and P.A.D., possible congruity effects are obscured by a generally higher level of con-

Figure 6. Semantic congruity index plotted separately for the set of stochastically dominant (unfilled circles) and the stochastically nondominant (filled circles) responses for the "easy" and the "difficult" pairs. The plots are based on pairs for which average response times were longer with the nondominant than with the dominant set of responses. The semantic congruity index is obtained by subtracting (mean of medians) response times with the instruction "easier" from average response times with the instruction "harder."

sistency in responding with the instruction "easier" than with the instruction "harder."

Intertask Response Time Relations

The analyses presented in this section were conducted with two purposes in mind. First, we sought to determine whether the requisite binary categorizations were indeed components of the comparisons of ease and of difficulty. Second, with a view toward establishing the properties of the time course of the comparative judgments of ease and difficulty, the possibility that the long-short categorizations were conducted in parallel with the binary comparisons of ease and of difficulty was examined by comparing, for each pair, the maximum binary categorization time with the corresponding comparison time. On the assumption that the two categorizations occur strictly in parallel with the binary comparison, comparison times cannot exceed the longer of the two categorizations, pro-

vided that the motor-output components are approximately equal. One model for how this could occur would assume that the component categorizations are initiated at the same time and that their durations till completion are monitored and compared. The first categorization to complete is the easier, and the slower is the harder.

Average categorization times with the stimulus with the slower categorization in each of the pairs adjacent in the I-Scale ordering are plotted at the bottom of Figure 4. As expected, categorization times monotonically increase as the pairs increase in difficulty of categorization. More importantly, as also can be seen in Figure 4, comparative judgments of ease and difficulty increase as the difficulty of the categorization with the component stimuli monotonically increases, and, as shown earlier, the linear trend is statistically reliable. It is also clear that comparative judgment times are uniformly longer than the categorizations. Thus, we can conclude that the requisite categorizations are components of the overall phenomenal comparative judgments of the ease and the difficulty of categorization. It is also clear that there is no evidence whatsoever for the strict parallel processing view. Most likely, the comparison process follows on the requisite binary categorizations.

Practice and the Form of the Latency Probability Function

Figure 7 plots mean (of medians with the various pairs) response times, conditional on the occurrence of a particular response, as a function of the probability of the response (i.e., LPFs) for each subject, for each third of the comparisons. For each block, points on the LPF were obtained by grouping the pairs with each instruction into the following frequency classes: {1,2,3}, {4,5,6}, {7,8,9}, {10, 11,12}, {13,14,15}, and {16}. Thus, for example, responses occurring 1, 2, or 3 times are stochastically non-dominant, and those occurring 13, 14, or 15 times in the block of 16 replications with each pair and each instruction are the corresponding stochastically dominant responses. The mean of the median response times for each pair with each instruction in each of the frequency classes was obtained and plotted against the midpoint of the frequency class divided by 16. Generally, for each subject,

Table 3
Response Probabilities for the Pairs Adjacent in the Ordering of Difficulty With Each Instruction

Subject	Instruction	Pair in the Difficulty Ranking							
		12	23	34	45	56	67	78	89
D.L.G.	harder	0.71	0.65	0.54	0.94	0.92	0.94	0.92	0.96
	easier	0.77	0.63	0.58	0.98	0.96	0.96	0.98	0.94
M.A.J.	harder	0.56	0.74	0.62	0.96	0.72	0.58	0.88	0.68
	easier	0.62	0.78	0.54	0.90	0.64	0.58	0.68	0.70
N.L.B.	harder	0.88	1.00	1.00	0.56	0.58	0.69	1.00	1.00
	easier	0.77	0.98	1.00	0.52	0.63	0.73	1.00	1.00
P.A.D.	harder	0.54	0.67	0.56	0.92	0.48	1.00	1.00	1.00
	easier	0.58	0.69	0.71	1.00	0.60	1.00	1.00	1.00

Note—The alternative labeled 1 is the most difficult and 9 the least difficult.

Figure 7. Latency probability functions (LPF) for each subject, as a function of practice. Diamond points are for the LPF based on the first block of trials (based on the first 16 replications with each pair with each instruction), circles for the LPF based on the second block, and triangles for the LPF based on the final block.

the LPF for the first block of comparisons is monotonically decreasing. However, as responses become faster with increasing practice, the LPFs do not remain monotonically decreasing. Rather, especially for Subjects M.A.J. and D.L.G., they become nonmonotonic for the second and third blocks of trials, thereby denying the occurrence of a single invariant form of LPF.

Separate analyses of variance, examining the linear and quadratic trend components with the response probability factor and the possible dependence of these components on block, were conducted for each subject. For Subject D.L.G., the LPFs exhibit a significant linear trend [$F(1,350) = 48.25, p < .0001$] and the response times significantly decrease over the three blocks [$F(2,350) = 18.61, p < .0001$]. In addition, the interaction of the linear trend with blocks approaches reliability [$F(2,350) = 2.66, p < .071$]. For Subject M.A.J., the LPFs exhibit reliable linear [$F(1,346) = 46.99, p < .0001$] and quadratic [$F(1,346) = 19.44, p < .0001$] trend components, and the main effect of block was also reliable [$F(2,346) = 6.33, p < .002$]. The change in the form of the LPF from monotonically decreasing on the first block to nonmonotonic on the subsequent blocks is evident in the form of the statistically reliable interaction of the linear trend component with block [$F(2,346) = 5.32, p < .005$]. Neither the main effect of block nor interactions involving block attained reliability for Subject N.L.B., and overall, both linear [$F(1,270) = 77.30, p < .0001$] and quadratic [$F(1,270) = 13.37, p < .0003$]

trends are evident. Finally, for Subject P.A.D., both linear [$F(1,288) = 66.07, p < .0001$] and quadratic [$F(1,288) = 5.33, p < .021$] trends are reliable, as is the main effect of block [$F(2,288) = 17.54, p < .0001$], but these trends remain unchanged over blocks; the interaction between blocks and trend components is not reliable.

GENERAL DISCUSSION

In contrast to the Petrusic and Jamieson (1989) study that required subjects to indicate which of two brightness patches was easier to distinguish, the present study forced the generation of component elements for the phenomenal comparison task. As indicated earlier, subjects in the Petrusic and Jamieson (1989) study could perform the comparison of ease of sameness-difference by rendering a dissimilarity comparison without engaging in the two implicitly required sameness-difference judgments on each trial. However, the evidence from the present study indicates that each of our 4 subjects was able to perform both the comparisons of ease and the comparisons of the difficulty of categorization in a meaningful way. In particular, because WST was satisfied and the obtained ordering was meaningful and appropriate for the task (e.g., the most difficult categorization in the comparison task was that with the alternatives nearest the point of subjective equality; i.e., the category boundary in the categorization task), we can infer that the overt phenomenal ease comparisons were based on the requisite binary categori-

zations of extent. Furthermore, the fact that phenomenal comparison times monotonically increased as the difficulty and, concomitantly, the time of the binary categorization increased provides further, converging, evidence that the requisite categorizations were performed.

At the outset, we indicated that Coombs's (1952, 1964) abstract data theory enabled us to map these phenomenal comparisons into preferential choice data and, in particular, to represent them in terms of his unidimensional theory of unfolding. Indeed, strong support for this view is obtained from the fact that the stochastically dominant orderings for each subject arise from folding the *a priori* J-Scale at the point corresponding to the "50% point" on the psychometric function obtained from the categorization task. Furthermore, since choices were replicated, the probabilistic versions of his unfolding theory were appropriate (e.g., Coombs, 1958; Coombs et al., 1959), and, accordingly, the selective effects of laterality on stochastic transitivity should be evident. Indeed, for the bilateral-split triples, SST was satisfied almost always, and SST was rarely satisfied with the bilateral-adjacent triples, precisely as expected. In addition, the present study establishes that comparisons are more difficult with bilateral pairs than with unilateral pairs. (See Coombs & Avrunin, 1988, chap. 4, for a discussion of the implications of this finding in terms of the difficulty of resolving a conflict.) Obtaining reliably longer decision times with the bilateral pairs provides further support for the view that subjects were indeed measuring the distance of the representation of each stimulus from the ideal point and then comparing these distances, as expected on the basis of Coombs's unfolding theory of preferential choice.

Reference point theories and semantic congruity. In order to account for semantic congruity effects, Jamieson and Petrusic (1975a) extended the measurement ideas of Coombs's unfolding theory of preference to develop their discrepancy ratio version of the general class of reference point theories (see also Holyoak, 1978, and Marks, 1972). Given the occurrence of semantic congruity effects in the present data, it is natural to apply the discrepancy ratio theory to provide a theoretical account of these comparative judgments of phenomenal ease and difficulty. According to this view, when subjects are instructed to select the alternative that is "more difficult" to categorize, they execute each categorization and then measure the distance of the representation of each stimulus from the category boundary; the stimulus *nearer* the category boundary is the more difficult. The decision of whether stimulus x or y is nearer to the category boundary, or referent, I_0 , is based on the ratio of the distances from the representation of each stimulus to the referent, I_0 , relative to a criterion, C , which in the typical unbiased case is set at 1. Jamieson and Petrusic (1975a) assume that response time is a monotonically decreasing function of the distance of the ratio $d(x, I_0)/d(y, I_0)$ from the criterion, C . When subjects are instructed to select the alternative that is "easier" to categorize, we assume that subjects select the alternative that is *farther* from the category boundary on the

folded J-Scale by measuring distances from the reference point, I^0 (for all x and y , $I_0 < x < y < I^0$), and computing the ratio above with I^0 substituted for I_0 . Congruity effects arise because for any given pair, x and y , the ratio is closer to the criterion, $C = 1$, when the pair is farther from the reference point than when it is closer; discriminability is reduced and response times increase as the distance between the stimulus pair and the activated reference point increase.⁸

The Jamieson and Petrusic (1975a) reference point theory applied in the context of the probabilistic version of Coombs's unfolding theory and embedded in an evidence accrual theory is able to provide a complete account of the present findings.⁹ The changing form of the LPF is predicted by the slow- and fast-guessing discrete state accrual model. We assume on each evidence accrual, the measurement of distances from the activated reference point and comparison to the criterion. The outcome of each of these underlying reference-point-based comparisons is then accrued in one of three discrete state accumulators, and the overt response is triggered when one of these accumulators reaches a preset criterion. The slow- and fast-guessing theory departs from the usual evidence accumulation models (see Luce, 1986; Pike, 1968; Townsend & Ashby, 1983; and Vickers, 1979, for models assuming the accrual of information in discrete time) by postulating the accrual of information that favors the occurrence of neither overt comparative judgment. Rather, information favoring the occurrence of overt indifference is also accrued. When the criterion for this indifference state is reached, a guessing response is emitted. The LPFs shown in Figure 7 suggest that during the early block of trials, the criterion for guessing remains relatively high and responding from the guessing state rarely occurs. In subsequent blocks, however, the criterion for guessing is lowered and the nonmonotonic LPFs observed arise because of an increase in the number of relatively fast guesses. As well, because the LPFs were neither flat nor strictly monotonically decreasing over all blocks, the response time predictions of the EBA and SOA theories are clearly disconfirmed. Of course, given the strong support for the unidimensional unfolding view of the present metacognitive judgments, the EBA and SOA theories are not appropriate because they apply primarily when the stimuli are multifeatured.

Semantic coding theory and the phenomenology of the task. It is natural to assume that subjects comply directly with the instructions on each trial; after each of the two categorizations, each categorization is then *further semantically coded* as "easy" or "hard." In a subsequent stage, these phenomenally based codes are then interrogated to permit the required binary comparison of phenomenal ease and difficulty. The outcomes from this experiment provide evidence against the view that the phenomenology of the task provides a viable model of these metacognitive-psychophysical comparisons. Although the phenomenal view might admit an ordering of the stimuli from least to most difficult to categorize, it

cannot accommodate the effects of laterality on stochastic transitivity and on response time.

The occurrence of congruity effects with ease and difficulty comparisons also suggests that the pairs are coded in terms of "ease" and "difficulty," and thus, indirectly, the congruity effect data also provide support for the intuitive categorization model. The propositionally based *semantic coding theory*, developed by Banks and his associates (e.g., Banks, 1977; Banks, Clarke, & Lucy, 1975; Banks, Fujii, & Kayra-Stuart, 1976, and more recently, Čech & Shoben, 1985; Čech, Shoben, & Love, 1990; Shoben, Čech, Schwanenflugel, & Sailor, 1989), is in fact precisely such a theory, which thus provides a potential "model" for these phenomenal ease and difficulty comparisons. Although the semantic coding model has provided effective first-order accounts of the available data, especially with symbolic comparisons, recently its limitations have become evident with numerical comparisons (Dehaene, 1989; Jaffe-Katz, Budescu, & Wallsten, 1989) and with perceptual comparisons (Petrusic, in press; Petrusic & Baranski, 1989a, 1989b), and in each of the reports above, the class of reference point theories has provided the more complete account.

The response time data also seriously question whether the semantic coding theory is appropriate for these phenomenal (formally preferential choice) comparisons. First, the theory is not able to account for the amplified congruity effects obtained with the more difficult bilateral comparisons. Second, in its current form, the semantic coding theory is not able to predict differential congruity effects as a function of stochastic dominance. With pair location and, consequently, semantic coding fixed (as is the case when dominance and nondominance are examined), the semantic coding theory has no mechanism other than the failures of code translation to modulate the magnitude of the congruity effect. As the result provided in the Appendix shows, the semantic coding theory predicts that congruity effects can never be larger with the set of stochastically nondominant responses than with the set of stochastically dominant responses, which is clearly contrary to the findings reported here. In psychophysical contexts, typically, stochastically dominant responses correspond to correct responses,¹⁰ and as Petrusic (in press) and Petrusic and Baranski (1989a, 1989b) have shown, congruity effects are never larger with correct responses than they are with error responses when error times exceed correct times. Thus, the semantic coding theory is unable to provide a complete account of the properties of the congruity effect in the present context of metacognition and in the broader context of perceptual comparison.

SUMMARY AND CONCLUSIONS

Requiring our subjects to make judgments about aspects of judgments rendered extends the domain of psychophysical judgment to include self-generated information. As such, by definition, we have provided a preliminary examination of metacognition in a rudimentary psychophysical task—the binary categorization of extent. In ad-

dition, the theoretical and experimental findings with phenomenal ease and difficulty comparisons provided in this article extend and complement the use of subjective measures in the assessment of mental workload (see Gopher & Donchin, 1986, and O'Donnell & Eggmeier, 1986). In the present case, it is evident that the use of comparative judgments of the subjective ease and the difficulty of a mental task permits a well-defined metric representation of mental workload.

Although binary categorization is an extremely simple task, the theoretical underpinnings permitting a representation of mental workload in binary categorization and a full understanding of the process of comparison of the ease and the difficulty of binary categorization are not. As we indicated earlier, a complete theory for the present admittedly simple problem must address a variety of separate issues, which, when taken together, have necessitated a rather elaborate and complex theoretical structure. Successful use of subjective measurement of mental workload in more "interesting" and "complex" tasks will necessarily require even more complex concatenations of theory than that developed here.

REFERENCES

- AUDLEY, R. J. (1960). A stochastic model for individual choice behavior. *Psychological Review*, 67, 1-15.
- AUDLEY, R. J., & PIKE, R. (1965). Some alternative stochastic models of choice. *British Journal of Mathematical & Statistical Psychology*, 18, 207-225.
- 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., CLARKE, H. H., & LUCY, P. (1975). The locus of the semantic congruity effect in symbolic comparisons. *Journal of Experimental Psychology: Human Perception & Performance*, 1, 35-47.
- 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*, 3, 435-447.
- BECHTEL, G. G. (1968). Folded and unfolded scaling from preferential paired comparisons. *Journal of Mathematical Psychology*, 5, 333-357.
- BIRNBAUM, M. H., & JOURNAL, J.-W. (1990). A theory of comparative response times and "difference" judgements. *Cognitive Psychology*, 22, 184-210.
- BUSEMEYER, J. R., FORSYTH, B., & NOZAWA, G. (1988). Comparisons of elimination by aspects and suppression of aspects choice models based on choice response-time. *Journal of Mathematical Psychology*, 32, 341-349.
- CARTWRIGHT, D. (1941). The relation of decision-time to the categories of response. *American Journal of Psychology*, 54, 174-196.
- ČECH, C. G., & SHOBN, E. J. (1985). Context effects in symbolic magnitude comparisons. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 11, 299-315.
- ČECH, C. G., SHOBN, E. J., & LOVE, M. (1990). Multiple congruity effects in judgments of magnitude. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 16, 1142-1152.
- COOMBS, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, 57, 145-158.
- COOMBS, C. H. (1952). *A theory of psychological scaling* (Engineering Research Institute Bulletin No. 34). Ann Arbor, MI: University of Michigan Press.

- COOMBS, C. H. (1958). On the use of inconsistency of preferences in psychological measurement. *Journal of Experimental Psychology*, **58**, 1-7.
- COOMBS, C. H. (1964). *A theory of data*. New York: Wiley.
- COOMBS, C. H., & AVRUNIN, G. S. (1988). *The structure of conflict*. Hillsdale, NJ: Erlbaum.
- COOMBS, C. H., GREENBERG, M., & ZINNES, J. L. (1959). A double law of comparative judgment for the analysis of preferential choice and similarities data. *Psychometrika*, **26**, 165-171.
- DEHAENE, S. (1989). The psychophysics of numerical comparison: A reexamination of apparently incompatible data. *Perception & Psychophysics*, **45**, 557-566.
- GOPHER, D., & DONCHIN, E. (1986). Workload—An examination of the concept. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance* (chap. 41, pp. 1-49). New York: Wiley.
- HALL, R. F. (1971). Laterality effects in the method of triads. *Perception & Psychophysics*, **10**, 101-103.
- HALL, R. F., & WEIR, R. (1974). Laterality effects in risk preference: A test of portfolio theory. *Acta Psychologica*, **38**, 351-355.
- HOLYOAK, K. J. (1978). Comparative reference points with numerical reference points. *Cognitive Psychology*, **10**, 203-243.
- JAMIESON, D. G., & PETRUSIC, W. M. (1975a). Relational judgments with remembered stimuli. *Perception & Psychophysics*, **18**, 373-378.
- JAMIESON, D. G., & PETRUSIC, W. M. (1975b). Presentation order effects in duration discrimination. *Perception & Psychophysics*, **17**, 197-202.
- JOHNSON, D. M. (1939). Confidence and speed in two-category judgment. *Archives of Psychology*, **34**, 1-53.
- JAFFE-KATZ, A., BUDESCU, D. V., & WALLSTEN, T. S. (1989). Timed magnitude comparisons of numerical and nonnumerical expressions of uncertainty. *Memory & Cognition*, **17**, 249-264.
- KRANTZ, P. H., LUCE, R. D., SUPPES, P., & TVERSKY, A. (1971). *Foundations of measurement*, (Vol. 1, pp. 145-150). New York: Academic Press.
- LINK, S. W. (1990). Modeling imageless thought. *Journal of Mathematical Psychology*, **34**, 2-41.
- LINK, S. W., & HEATH, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, **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.
- MORGAN, B. J. F., & ROBERTSON, C. (1980). Short-term memory models for choice behavior. *Journal of Mathematical Psychology*, **21**, 30-52.
- 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.
- O'DONNELL, R. D., & EGGMEIER, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance* (chap. 42, pp. 1-49). New York: Wiley.
- PETRUSIC, W. M. (in press). Semantic congruity effects and theories of the comparison process. *Journal of Experimental Psychology: Human Perception & Performance*.
- PETRUSIC, W. M., & BARANSKI, J. V. (1989a). Context, context shifts, and semantic congruity effects in comparative judgments. In D. Vickers & P. Smith (Eds.), *Human information processing: Measures, mechanisms and models* (pp. 231-251). Amsterdam: North-Holland.
- PETRUSIC, W. M., & BARANSKI, J. V. (1989b). Semantic congruity effects in perceptual comparisons. *Perception & Psychophysics*, **45**, 439-452.
- PETRUSIC, W. M., & BARANSKI, J. V. (1991). *On the internal psychophysics: Binary and quaternary relational judgments with remembered and perceived magnitudes*. Manuscript submitted for publication.
- PETRUSIC, W. M., COUSINS, L. S., & CORBIN, R. (1984). Rare event probabilities unfold. *Canadian Journal of Psychology*, **38**, 478-491.
- PETRUSIC, W. M., & JAMIESON, D. G. (1976). Studying preferential choice without stable probability estimates: Temporal transitivity analyses. *Acta Psychologica*, **40**, 375-383.
- 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. (1968). Latency and relative frequency of response in psychophysical discrimination. *British Journal of Mathematical & Statistical Psychology*, **21**, 161-182.
- RESTLE, F. (1961). *Psychology of judgment and choice*. New York: Wiley.
- SHIPLEY, W. C., COFFIN, J. I., & HADSELL, K. C. (1945). Reaction time in judgments of colour preferences. *Journal of Experimental Psychology*, **35**, 206-215.
- 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.
- SHOBEN, E. J., CECHE, C. G., SCHWANENFLUGEL, P. J., & SAILOR, K. M. (1989). Serial position effects in comparative judgments. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 273-286.
- SMITH, P. L., & VICKERS, D. (1988). The accumulator model of two-choice reaction. *Journal of Mathematical Psychology*, **32**, 135-168.
- TOWNSEND, J. T., & ASHBY, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
- TVERSKY, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, **79**, 281-299.
- VICKERS, D. (1979). *Decision processes in visual perception*. New York: Academic Press.
- VICKERS, D., CAUDREY, D., & WILLSON, R. J. (1971). Discriminating between frequency of occurrence of two alternative events. *Acta Psychologica*, **35**, 151-172.
- 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.

NOTES

1. These axioms impose severe constraints on the quaternary relation. For example, the transitivity condition requires that if it is easier to discriminate between a and b in the stimulus pair (a, b) than it is to discriminate between c and d in the pair (c, d) (denoted $abEcd$), and if it is easier to discriminate between c and d in the pair (c, d) than it is to discriminate between e and f in the pair (e, f) , then it must be the case that it is easier to discriminate between a and b in the pair (a, b) than it is to discriminate between e and f in the pair (e, f) . In addition, the monotonicity axiom provides an extremely exacting test of the requirement that intervals can be concatenated in an orderly manner. Monotonicity is defined thus: for all a, b, c, a', b' , and c' in the set of stimuli, if $abEa'b'$ and $bcEb'c'$, then $acEa'c'$.

2. Semantic congruity effects with perceptual comparisons were first reported by Audley and Wallis (1964), who used brightness comparisons, and by Wallis and Audley (1964), who used pitch comparisons. In addition to bringing the phenomenon to the attention of contemporary experimentalists, they also provided a quantitative account of the effect in the context of one of the earliest evidence accumulation models (see Audley, 1960). See Petrusic (in press) for further theoretical and empirical examination of the effect with perceptual comparisons and Petrusic and Baranski (1991) for studies of the effect with remembered comparisons.

3. We could have defined the stochastic dominance criterion with respect to each instruction; however, except for occasional nonsystematic variations, the stochastically dominant orderings coincide with the two instructions. Consequently, we define pairwise dominance after averaging over the two comparison instructions.

4. The rank orderings of difficulty in the comparison task are well predicted on the assumption that they are based on folding at the category boundary between "long" and "short" on the *perceptual* length scale. On the assumption that the psychophysical function for length is compressive and is given by a power function with exponent 0.82 (based on the values reported in Petrusic & Jamieson, 1979), distances from the ideal point to each stimulus on this scale were computed. The precise location of the ideal point between the two most difficult stimuli in the ordering was obtained by linear interpolation on the subjective (perceptual) scale, on the basis of the proportion of times that the stimulus ranked as most difficult was judged to be above the stimulus that was next most difficult. The orderings obtained on this basis were: 879654321, 657483921, 567438291, 675894321, for Subjects D.L.G., M.A.J., N.L.B., and P.A.D., respectively. For Subjects D.L.G., M.A.J., and P.A.D., the orderings with the bilateral pairs are well predicted on this basis. Not surprisingly, nearly identical rankings are predicted if it is assumed that subjective magnitudes are linearly related to physical extent. Nevertheless, prediction of the ordering of subjective difficulty is enhanced (albeit, slightly) if the psychophysical function is compressive. On the other hand, for Subject N.L.B., prediction with an assumed exponent of 0.82 is not so good. It is improved somewhat if the function is made even more compressive. The ordering obtained on the basis of a *logarithmic* psychophysical function is 564789321. Petrusic, Cousins, and Corbin (1984) should be consulted for a further example of how inferences concerning the metric on the J-Scale, and, consequently, the form of the psychophysical function can be based on the particular subset of I-Scales occurring. We are grateful to an anonymous reviewer for suggesting the preceding analyses.

5. Several tests of SST with the unilateral triples may have been inappropriate because the response probabilities for each of the pairs in the triad were often one or very close to one (e.g., the response probabilities for the embedded pairs are 0.99 and 1.00, and the probability for the bounding pair in the triple is 0.99). If triples involving such pairs are excluded, SST is rarely violated with the unilateral triples.

6. The main effect of instruction was not uniform over the 4 subjects: failing to attain significance, for Subject P.A.D.; significantly faster responding with the instruction "harder," for Subject M.A.J. [$F(1,40) = 4.14, p < .049$]; and significantly faster responding with the instruction "easier," for the two remaining subjects [$F(1,38) = 23.79, p < .001$, and $F(1,32) = 16.31, p < .0001$, for Subjects N.L.B. and D.L.G., respectively].

7. If, in order to conduct the analyses with greater statistical power, median response times with each pair and instruction for each session are also viewed as replicates in the 8 (pairs) $\times$ 2 (instructions) completely within-subject factorial design, then statistically reliable congruity effects are evident in the interaction between the linear trend with the location of the pairs on the length continuum and the instruction [$F(1,23) = 16.05, p < .0006$].

8. Petrusic (in press, Experiments 1 and 2), Petrusic and Baranski (1989a), and Petrusic and Baranski (1989b, Experiment 3) required subjects to select the point on either side of a central fixation point that was either "nearer to" or "farther from" the central fixation point. It is interesting that the formal representation of this task is identical to that with the bilateral pairs in the ease and the difficulty comparisons reported here. In the ease and the difficulty comparison task, the category boundary, I_0 , which also serves as the referent, corresponds to the central fixation point in the perceptual proximity/distance task. It differs, however, in that in the visual proximity comparison tasks, it is perceptually available, but in the phenomenal ease and difficulty comparison task, it is mentally constructed.

9. In order to account for the failure to observe clear semantic congruity effects with the response probability measure, the level of complexity of a full explanation will have to be increased further. It also may be necessary to assume that subjects adjust the evidence accrual criteria to maintain an approximately equal level of response consistency

with the two forms of instruction (cf. Petrusic, in press; Petrusic & Baranski, 1989a, 1989b).

10. In the majority of tasks, the correct response is indeed the stochastically dominant response. However, in duration discrimination (e.g., Jamieson & Petrusic, 1975b), where presentation order effects can be sizable, the stochastically dominant response can be incorrect. Nevertheless, in this case, the stochastically dominant error response exhibits the properties of correct responses, typically found in other tasks, where they are stochastically dominant (e.g., stochastically dominant response times are faster than stochastically nondominant response times when subjects respond cautiously, sacrificing speed for accuracy).

APPENDIX

In this section, we will obtain an easily testable property of the discrete semantic coding (Banks, 1977) and the response translation (Wallis & Audley, 1964) theories of the semantic congruity effect. We will develop expressions for the magnitude of the congruity effect and show that congruity effects can never be larger for the set of stochastically nondominant responses than for the set of stochastically dominant responses.

Let R_1 and R_2 denote Responses 1 and 2 and I_1 and I_2 , Instructions 1 and 2, respectively. For notational convenience, suppose code translation is not required with Instruction 1 and is with Instruction 2. With all other factors fixed, differential congruity effects can arise as a consequence of failures to translate either codes or responses when required. Let t denote the probability of translation when required and let c be a constant, the time for code translation. When code translation is not required, $P(R_1|I_1) = p$ and $P(R_2|I_1) = (1-p) = q$. When translation is required, R_1 responses occur with probability $P(R_1|I_2) = pt + q(1-t)$ and R_2 responses have probability $P(R_2|I_2) = qt + p(1-t)$. For $p > q$, whenever $t > 0$, semantic congruity effects are evident with the stochastically dominant response; that is, $p > pt + q(1-t)$.

Average response times with Instruction 1 are given by $T(R_1|I_1) = T_1$ and $T(R_2|I_1) = T_2$. However, when code translation is required, $T(R_1|I_2) = [pt(T_1+c) + q(1-t)T_1]/[pt+q(1-t)]$, which simplifies to $T(R_1|I_2) = T_1 + ptc/[pt+q(1-t)]$. Similarly, $T(R_2|I_2) = T_2 + qtc/[qt+p(1-t)]$. We define the semantic congruity index for each response, $SCI(R_i) = T(R_i|I_2) - T(R_i|I_1)$. If $p > q$, R_1 is the stochastically dominant response and $SCI(R_1) = ptc/[pt+q(1-t)]$, and, for the stochastically nondominant response, $SCI(R_2) = qtc/[qt+p(1-t)]$. Finally, the difference in semantic congruity magnitude between the stochastically dominant and the stochastically nondominant responses is given by

$$SCI(R_1) - SCI(R_2) = t(1-t)c(p^2 - q^2)/[pt+q(1-t)][qt+p(1-t)]. \quad (A1)$$

Since $p > q$, this expression is positive whenever both t and $c > 0$; consequently, congruity magnitude is never larger with the stochastically nondominant response than it is with the stochastically dominant response according to the semantic coding and probabilistic response translation theories.