

# Mixed-method mixed-modality psychophysical scaling

LAWRENCE M. WARD

*University of British Columbia, Vancouver, British Columbia*

Four experiments are reported in which the direct psychophysical scaling methods of magnitude estimation, category judgment, and cross-modality matching were mixed in the same series of trials, both with a single stimulus modality and in a mixed-modality situation. The mixed-method scaling situation gave results consistent with those obtained when methods are used alone, and it has several advantages. Interactions between the methods were consistent with the idea that judgments made under all three are mediated by a primitive process of categorization that is influenced by heuristics used to achieve a single category identity for each stimulus.

In the roughly 130-year history of psychophysics, many ingenious techniques have been invented with which the magnitude of sensory experience has been assessed. Ideally, the results found with all of these techniques should converge on reliable and valid scales for the measurement of the magnitude of sensory experience in various sensory modalities. Unfortunately, however, such convergence has not been the case, and controversy continues as to which scales are valid, and even as to what the validity criteria should be. Thus, further insight into the properties of psychophysical scaling methods is needed.

The present paper focuses on three methods that are usually considered to be "direct" scaling methods: magnitude estimation, category judgment, and cross-modality matching. All three are commonly used, and separately, they have given rise to many useful measurements of sensory experience. Also, they do generate a reasonable degree of scale convergence (see, e.g., Krueger, 1989; S. S. Stevens, 1975). Yet these methods are not fully understood *qua* judgment processes. That is, *why* convergence is or is not obtained is not predictable from a validated general model of psychophysical scaling judgment, although several have been proposed (e.g., those of Lockhead & King, 1983; Luce, Baird, Green, & Smith, 1980; Marley & Cook, 1986; Treisman & Williams, 1984; Ward, 1979). The usual approach to studying these methods involves having observers make judgments of sensory magnitude of stimuli from a single modality, using one particular method. Several such studies done with different modalities of stimuli and/or different judgment methods are then compared and general conclusions sought (see, e.g., Ward, 1987). A fruitful alternative has been the mixed-modality scaling approach (Ward, 1982, 1985,

1986), which was introduced by J. C. Stevens and Marks' (1980) work in magnitude matching and anticipated by earlier work in cross-continuum anchoring (e.g., Aftanas & Rule, 1968). In this approach, stimuli from different modalities are presented in various sequences for judgment by the same method, resulting in indirect cross-modality matching functions generated by magnitude estimations, category judgments, and cross-modality matches of "difficult" stimulus continua to an "easier" response continuum. The data from such experiments also bear on theories of psychophysical judgment. For example, I have established that in this context, contrastive dependencies of current responses on previous stimuli are usually confined to previous stimuli of the same modality, whereas assimilative effects of previous responses do not depend on the modality of the stimulus to which they were made. This finding effectively rules out theories in which both types of sequential effects arise from the same mechanism.

The present paper extends the "mixing" approach to judgment methods as well as stimulus modalities. Thus, I report here the results of four experiments in which observers made judgments of sensory magnitude according to two different methods interspersed in the same sequence of judgment trials. In the first two experiments, only methods were mixed. Observers judged either the brightness of lights or the loudness of sounds, using an alternating sequence of judgment methods. In Experiment 1, magnitude estimations alternated with cross-modality matches (using a duration response; see Ward, 1986), as in the sequence

$ME_1 CMM_1 ME_2 CMM_2 ME_3 CMM_3, \dots$

whereas in Experiment 2, category judgments alternated with cross-modality matches, as in the sequence

$CJ_1 CMM_1 CJ_2 CMM_2 CJ_3 CMM_3, \dots$

No experiments were done in which magnitude estimations and category judgments were mixed, since both require a numerical response and it was thought that less useful information would result from a mixture of such

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. I thank Odeis Geiger for help with running observers and analyzing data. Address reprint requests to Lawrence M. Ward, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC V6T 1Y7, Canada.

similar methods. However, both of those methods have been considered to be quite different from cross-modality matching (e.g., S. S. Stevens, 1975), even though similar sequential dependencies and long-term memory effects have been found in all three (Ward, 1987). Thus there is reason to believe that responses on these methods would interact in an interesting way, although the differences in response modality (numbers versus, in this case, durations) and the different instructions given observers make it difficult to make quantitative predictions about these interactions without making several arbitrary assumptions.

However, one set of semiquantitative predictions can be made from the fuzzy judgment theory of psychophysical judgment (Ward, 1979). Briefly, this theory proposes that for each stimulus presented to an observer in a judgment task, an internal representation will be formed on the basis of the activity of the sensory system that has transduced the stimulus. The internal representation is assumed to be a fuzzy subset (Zadeh, 1965) of the set of possible sensation levels, and to have an excitatory center-inhibitory surround organization that results in mutual repulsion (contrast) of internal representations on current and previous trials (cf. Levine & Grossberg, 1976). The internal representation (after repulsion) is then cross-correlated with each of the prototypes (also fuzzy subsets) of categories stored in long-term memory and rank ordered with respect to their central tendencies. This (perceptual) process often gives rise to more than one candidate category (satisfactory match). The resulting uncertainty is resolved by a heuristic process that chooses the category nearest to that used previously. This causes assimilation of the current category to the previous category, and this is passed on to the response stage through mapping of the chosen stimulus category to an equivalent category on the response continuum, followed by choice of a response whose magnitude sufficiently matches the prototype of that response category.

In the fuzzy judgment approach, the process of categorization is assumed to be central to all of the so-called "direct" judgment methods. Differences between methods arise from constraints placed by the experimenter on the way responses are generated, in terms of both the frame of reference and the specific response set allowed (Marks, Szczesuil, & Ohlott, 1986; Ward, 1987). Dependencies of current responses on previous responses arise from heuristics used in the categorization process to choose a single category from an "indifference set" of plausible candidates. Thus, mixtures of two methods should give rise to dependencies of responses in one method on those of the other method, regardless of differences in the response output processes, since the dependencies arise from a process common to both judgment methods. The relative magnitudes of these effects, as measured by linear regression (see later) should depend on the particular stimulus and response sets in use. Other approaches that emphasize differences in the fundamental processes through which responses are arrived at in the several judgment methods would predict no such effects. Dependen-

cies of current responses on previous stimuli should also be independent of judgment method, since in the theory they arise from contrastive sensory processes that are the same in all methods; consistent with previous results, such effects should be contrastive and should be much larger when the previous stimulus is from the same modality than when it is from a different modality.

Experiments 3 and 4 combine the mixed-method approach introduced in Experiments 1 and 2 with the mixed-modality approach studied earlier. In these experiments, stimulus modality was alternated by pairs (e.g., two successive judgments of sounds, S, followed by two successive judgments of lights, L, etc.) while judgment methods alternated strictly, giving rise in Experiment 3 to the sequence

ME-S<sub>1</sub> CMM-S<sub>2</sub> ME-L<sub>1</sub> CMM-L<sub>2</sub> ME-S<sub>3</sub> CMM-S<sub>4</sub>  
ME-L<sub>3</sub> CMM-L<sub>4</sub>, . . . ,

and in Experiment 4 to the sequence

CJ-S<sub>1</sub> CMM-S<sub>2</sub> CJ-L<sub>1</sub> CMM-L<sub>2</sub> CJ-S<sub>3</sub> CMM-S<sub>4</sub> CJ-L<sub>3</sub>  
CMM-L<sub>4</sub>, . . .

In these experiments, it was possible to study the effects of both method and modality on sequential dependencies of the current response on previous stimuli and responses.

These experiments may be seen as valuable beyond the light they may throw on theories of psychophysical judgment. For example, in clinical and laboratory tests, several different methods of psychophysical judgment are often used successively (although not in the close proximity used here). If experimenters hold the view that very different methods do not interact, the contamination of results by unexpected interactions could go unobserved. Also, issues concerning the equivalence of response modalities, or the comparability of scaling methods (cf. Krueger, 1989) are addressed directly by such studies. It is even possible to compare indirectly derived scales, as, for example, those of duration in the present experiments, with those obtained directly, in order to check for scale convergence that would validate both approaches.

## METHOD

### Experiment 1: ME and CMM

**Observers.** Four observers with normal or corrected-to-normal vision judged lights, and 4 observers with no known hearing defects judged sounds. All observers were paid for participating. Seven were males and 1 was female. All were undergraduates at the University of British Columbia with no previous experience in experiments of this sort.

**Apparatus.** The apparatus used was identical to that described by Ward (1982). Briefly, in the present experiment a (pseudo) random sequence of sounds or lights, varying in intensity, was presented for either magnitude estimation or cross-modality matching judgment. The light stimuli consisted of a 1.5° dot of approximately uniform luminance at eye level for the observer, who sat in a model 250A IAC sound-attenuating chamber. The 10 light intensities used ranged from 0.328 cd/m<sup>2</sup> to 168.0 cd/m<sup>2</sup> in 3-dB steps. Adjacent luminances were separated by 0.3 log units (3 dB). The 10 sound stimuli consisted of 1000-Hz tones, delivered diotically through stereo headphones, and ranging from 0.063 dynes/cm<sup>2</sup> to 11.25 dynes/cm<sup>2</sup> in

5-dB steps. Adjacent amplitudes were separated by 0.25 log units (5 dB). Each stimulus was presented for 1 sec. A DEC PDP-11/34A computer controlled the stimulus presentation and recorded the responses, which were made on a standard computer keyboard illuminated by a dim red light.

**Procedure.** All observers made a 200-trial practice run and four 300-trial data runs, for a total of 1,200 data trials—approximately 600 per judgment method and about 60 per intensity level within each method. Stimulus intensities varied (pseudo) randomly, and methods of judgment alternated strictly, as described in the introduction. The type of judgment (ME or CMM) to be made was indicated on each trial by illumination of one of two appropriately labeled LEDs. Magnitude estimation instructions were standard, with a modulus of 100 assigned to the 5.250-cd/m<sup>2</sup> or the 0.632-dyne/cm<sup>2</sup> stimulus as appropriate. For cross-modality matching, observers indicated a duration by pressing a particular key twice, with the presses separated by the desired duration response. This is a simple and reliable way to generate such responses (see Ward, 1986). Observers were dark-adapted for about 2 min before judging the light stimuli. However, since intertrial intervals seldom exceeded 10 sec, complete dark adaptation probably did not occur following the most intense light stimuli.

**Experiment 2: CJ and CMM**

**Observers.** Three different observers with normal or corrected-to-normal vision judged lights, and 5 different observers with no known hearing defects judged sounds. All observers were paid for participation. Three were females and 5 were males. All were undergraduates at the University of British Columbia, and none had any previous experience in experiments of this sort.

**Apparatus.** The apparatus and stimuli were identical to those used in Experiment 1.

**Procedure.** The procedure was identical to that in Experiment 1, except that the observers were given category judgment instructions instead of magnitude estimation instructions. Before each run, the observers were presented with the least and most intense stimuli and asked to divide the intervening intensity range into 10 equal categories, labeled 1 to 10. They were to assign the least intense to Category 1, the most intense to Category 10, and the other intensities to Categories 1-10 as appropriate. Cross-modality matching responses (durations) were generated as in Experiment 1. Thus, in this experiment, category judgments alternated with cross-modality matches of duration to sound or light intensity.

**Experiment 3: ME and CMM and Mixed Modalities**

**Observers.** Eight different observers with normal or corrected-to-normal vision and no known hearing defects judged lights and sounds. All observers were paid for participation. Three were females and 5 were males. All were undergraduates at the University of British Columbia, and none had any previous experience in experiments of this sort.

**Apparatus.** The apparatus and stimuli were identical to those used in Experiment 1.

**Procedure.** The procedure was identical to that in Experiment 1, except that stimulus modality alternated by pairs as described in the introduction. Thus, in Experiment 3, observers made a magnitude estimation of a sound (ME-S<sub>1</sub>), followed by a cross-modality match of a duration to a sound (CMM-S<sub>2</sub>), followed by a magnitude estimation of a light (ME-L<sub>1</sub>), followed by a cross-modality match of a duration to a light (CMM-L<sub>2</sub>), and repeated this sequence 75 times per 300-trial run. Stimulus intensities were selected (pseudo) randomly, as before.

**Experiment 4: CJ and CMM and Mixed Modalities**

**Observers.** Six different observers with normal or corrected-to-normal vision and no known hearing defects judged lights and sounds. All observers were paid for participation. Four were females and 2 were males. All were undergraduates at the Univer-

sity of British Columbia, and none had any previous experience in experiments of this sort.

**Apparatus.** The apparatus and stimuli were identical to those used in Experiment 1.

**Procedure.** The procedure was identical to that in Experiment 3, except that category judgments were substituted for magnitude estimations.

**RESULTS AND DISCUSSION**

**Response Matching Functions**

In Experiments 1 and 2, observers judged the magnitudes of the same set of sounds or lights, using either of two different methods alternately. Therefore, it is possible to plot the average response given to a certain stimulus in one method against that given the same stimulus in the other method. Such plots are shown in Figure 1 for Experiments 1 and 2 for the data averaged across observers. These *response matching functions* display the relationship between the response continua (number for magnitude estimation and category judgment, duration for cross-modality matching) used in the various methods when applied to the same stimuli by the same observers.

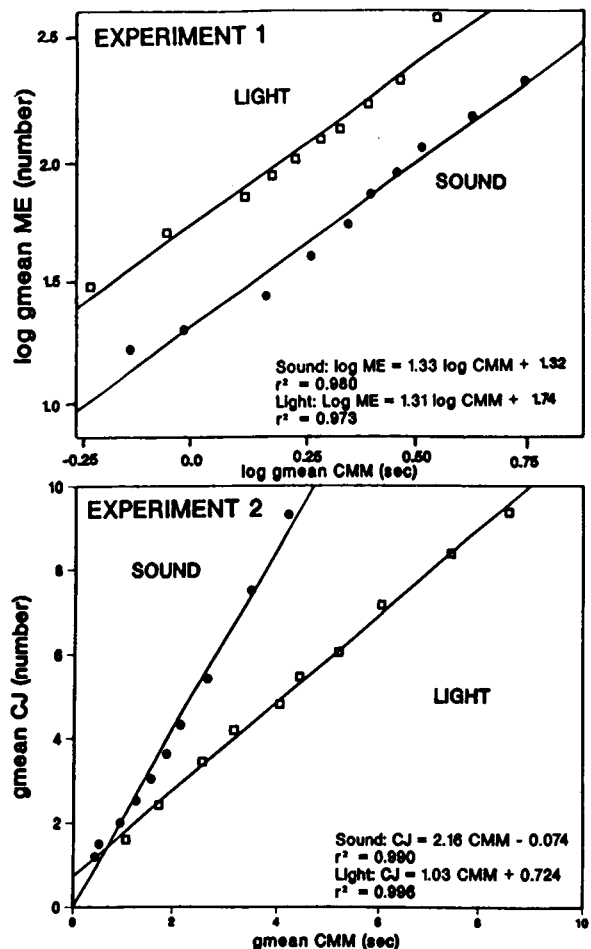


Figure 1. Response matching functions for Experiment 1 (magnitude estimation and cross-modality matching) and Experiment 2 (category judgment and cross-modality matching).

For the data of Experiment 1, mean log magnitude estimations are plotted against mean log cross-modality matching durations, since the relationships in the untransformed plots were highly curved. Geometric means are displayed for the data of Experiment 2 because of skewness; plots of arithmetic average responses are highly similar. (It should be noted that standard psychophysical functions for all data from the present experiments closely resemble those found when each method is used alone—cf. Ward, 1987—and are not reproduced here in order to conserve space.)

There are several noteworthy aspects to the data displayed in Figure 1. First, it is clear that there is a simple, though not consistent, relationship between the various response continua. This relationship varies with stimulus modality, although it must be kept in mind that different groups of observers generated each of the four matching functions shown. The functions for sound and light stimuli in Experiment 1 are reasonably approximated by power functions with exponents of about 1.3, somewhat larger than the value of approximately 1.0 usually obtained in direct matching of number to duration over the range of durations used by these observers.<sup>1</sup> The matching functions for Experiment 2 present a somewhat different picture. There the functions are approximately linear in untransformed plots, but the slopes and intercepts of the linear functions differ greatly. Again, since different groups of observers generated the two plots, it is possible that the differences in number and duration use between the groups arose incidentally rather than from a consequence of differences in method (re Experiment 1 functions) or stimulus modality. The discussion of the response matching functions from Experiments 3 and 4 lends more credibility to the idea that the differences observed here were caused by some unknown aspects of the experiments and/or the observer groups. However, this does not alter the conclusion that numbers generated as responses in the mixed-method context do not have the same relation to durations generated as responses to the same stimuli as they do when the numbers are supposed to directly reflect the experienced magnitude of a duration stimulus.

Figure 2 displays the response matching functions for the data of Experiments 3 and 4. These are generated from the trials on which the different methods were used to judge the magnitudes of stimulus intensities from the same modality (e.g., ME-S<sub>1</sub> and CMM-S<sub>2</sub>). Here the matching functions have the same simple character as do those in Figure 1, but they are much more similar within each panel. Here also, the relationship between number judgments and durations used as cross-modality matching responses is different from that obtained when numbers are used to judge durations directly. The slopes of the functions from Experiment 3 (1.82 and 1.57 for sound and light, respectively), which represent exponents of simple power functions, are larger than the exponent usually obtained from direct magnitude estimations of durations, and they are also larger than those obtained in Experi-

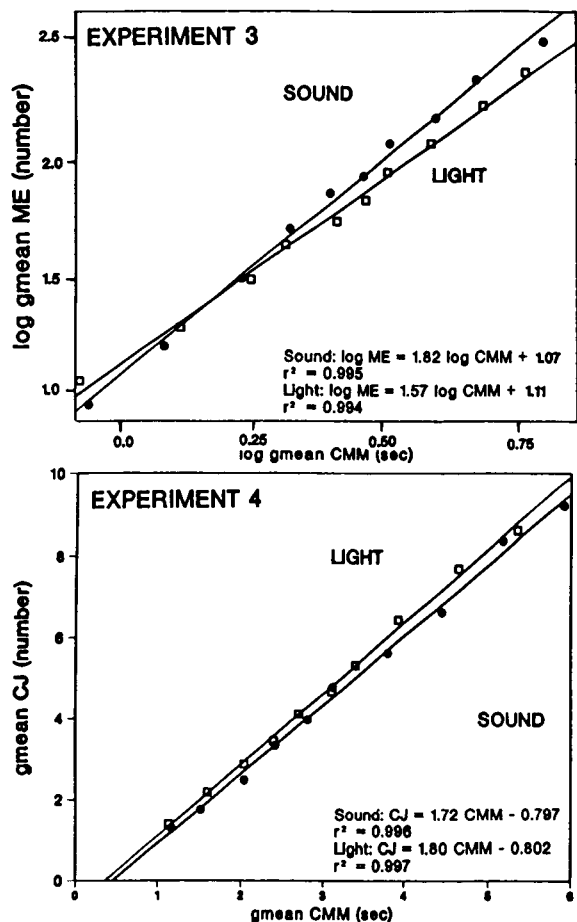


Figure 2. Response matching functions for Experiment 3 (magnitude estimation and cross-modality matching) and Experiment 4 (category judgment and cross-modality matching).

ment 1. The functions from Experiment 4, although reasonably linear in untransformed coordinates, have slopes of 1.72 and 1.80 for sound and light, respectively, indicating that category number and duration responses (in seconds) were not equivalent; in particular, category number was not simply an estimate of duration nor was duration simply a translation of category number into seconds.

When the same observers generate responses in the same series of trials to both light and sound stimuli, the resulting response matching functions are highly similar (Figure 2), whereas when different groups of observers generate responses to the two different modalities, the functions can be quite different (Figure 1). One explanation for this difference is that judging both modalities with both methods in the same session causes observers to attempt to use responses more consistently across modalities. Observers in Experiments 1 and 2, who judged either light or sound stimuli but not both, could have made responses more idiosyncratically, since there was no requirement to be consistent with the group of observers judging the "other" modality. This can be seen most

clearly in Figure 1 for the Experiment 2 data. The average response range for the category judgment method was about the same for both sound and light, but the range of durations used by the observers who judged sounds was about 0.4–4 sec, whereas that used by the observers who judged lights was about 1–8.5 sec. Why the two groups of observers chose to use different ranges of durations is unclear; that they chose to do so is reflected in the slope difference in the response matching functions. Similarly, observers who judged lights in Experiment 1 used about the same (log) ranges of numbers and durations as those who judged sounds, but the numbers they used were almost 0.5 log units higher on the average. Again, it is unclear why the different groups of observers chose different magnitudes of numbers; that they did is reflected in the intercept differences in the response matching functions. Clearly, the response matching functions shown in Figure 2 are to be preferred.

**Indirect Cross-Modality Matching Functions**

The observers in Experiments 1 and 2 judged only stimulus intensities on a single modality. Furthermore, the various groups of observers did not use their responses in the same way (see above). Therefore, it is not feasible

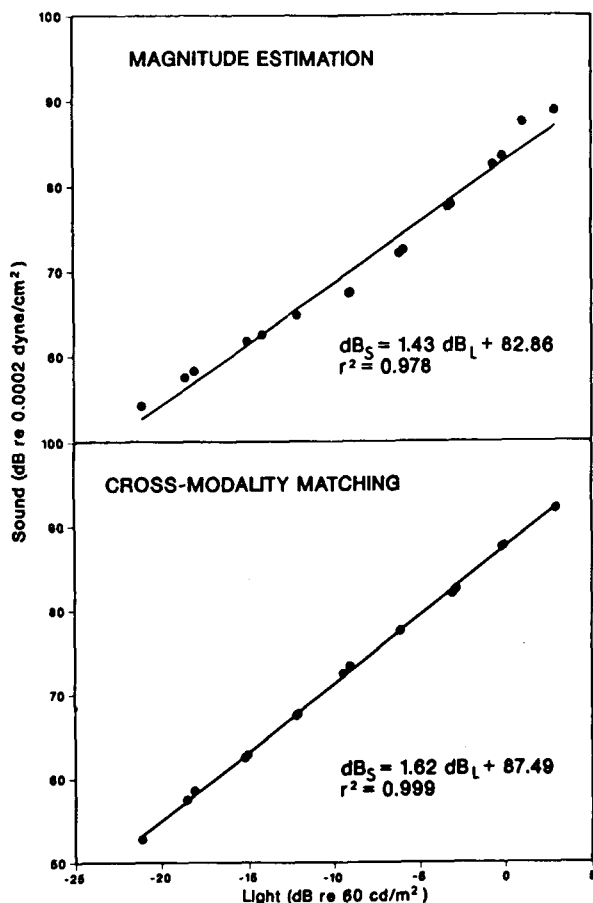


Figure 3. Derived cross-modality matching functions from Experiment 3.

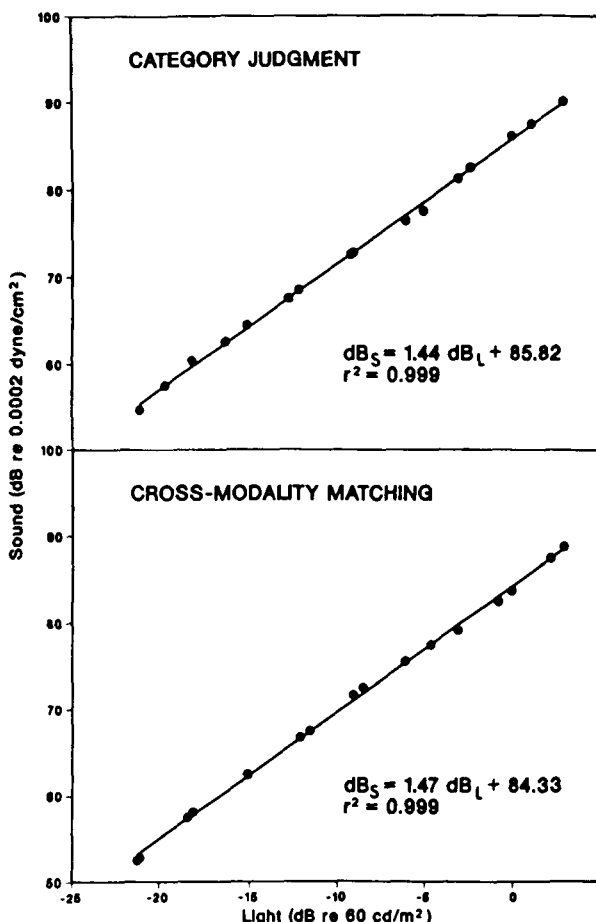


Figure 4. Derived cross-modality matching functions from Experiment 4.

to derive sound-intensity-light-intensity matching functions from the data of those experiments. However, the observers in Experiments 3 and 4 judged both sound and light intensities in the same series of trials and made judgments of each type of stimulus, using the same judgment method. In addition, they appear to have used their responses similarly for each modality (Figure 2). Therefore, it is feasible to construct indirect cross-modality matching functions for these data, using the procedure of J. C. Stevens and Marks (1980, Appendix 2). In this procedure, adjacent pairs of stimulus intensities (in decibels) and log geometric mean responses are each averaged to produce smoother psychophysical functions. Then linear interpolation is used to determine for each of the (averaged adjacent) intensities on one modality which intensity on the other modality would have been given the same response. This procedure is then repeated for the "other" modality with respect to the first. The resulting sets of derived "matches" make up the indirect cross-modality matching functions.

Figures 3 and 4 show the indirect cross-modality matching functions obtained in this way from the data of Experiments 3 and 4, respectively. Each function is derived

from the geometric mean responses on the appropriate trials across all observers in a given experiment. For example, the magnitude estimation function in Figure 3 is based on the ME-S<sub>1</sub> and ME-L<sub>1</sub> judgments in the sequence of trials of Experiment 3. Functions for individual observers resemble these average functions quite closely. Clearly, these data produce excellent fits to straight lines in log-log (dB-dB) plots, actually somewhat better fits than the more standard mixed-modality method does, both visually (substantially less nonlinearity) and in terms of *r*<sup>2</sup> (here *r*<sup>2</sup> = 0.999 in three cases, *r*<sup>2</sup> = 0.978 in the other; *r*<sup>2</sup> = from 0.975 to 0.991 in Ward, 1982, 1985, 1986).

In Figures 3 and 4, only the sound-on-light regression equations are displayed. Comparable regression equations from previous experiments (with the same general situation, but with only mixed-modality scaling along with a single judgment method) had the following slopes for the strict alternation and the alternation-by-pairs stimulus sequences, respectively: magnitude matching, 1.59 (Ward, 1982) and 1.68 (Ward, 1985); category matching, 1.45 (Ward, 1982) and 1.59 (Ward, 1985); double cross-modality matching, 1.70 and 1.63 (Ward, 1986). These slopes compare favorably with the present values of 1.43, 1.44, and 1.62 and 1.47 for magnitude matching, category matching, and double cross-modality matching, respectively (see Figures 3 and 4). Apparently mixing methods does not greatly alter these derived cross-modality matching functions and may even have the effect of producing cleaner data. This may be because the mixing of judgment methods and stimulus continua makes the judgment sessions more interesting; informal impressions of observers' reactions to the situation support this view.

**Microanalysis of Response Dependencies**

**Multiple regression analyses.** Sequential dependencies of the current response on previous stimuli and responses

have been studied in a variety of psychophysical scaling situations, including mixed-modality scaling (e.g., Ward, 1982, 1985, 1986). One useful way to describe these effects has been to estimate  $\gamma$ ,  $\alpha_i$ , and  $\beta_k$  in the following regression equation:

$$\log R_n = \gamma \log I_n + \sum_{i=1}^M \alpha_i \log I_{n-i} + \sum_{k=1}^N \beta_k \log R_{n-k} + \delta + \epsilon \quad (1)$$

where *R<sub>n</sub>* is the response on trial *n*, and *I<sub>n</sub>* is the stimulus intensity on trial *n* (see Jesteadt, Luce, & Green, 1977; Ward, 1979). For the present data, these regression coefficients were estimated for the data of individual observers by hierarchical linear regression, in which the independent variables were entered into the regression equation in the order of the most "recent" (closest to *R<sub>n</sub>*), next most recent, and so forth (e.g., *I<sub>n</sub>* first, *R<sub>n-1</sub>* next, then *I<sub>n-1</sub>*, etc.). For the coefficients reported here, *M* = 4 and *N* = 4, since coefficients for higher lags are seldom reliable. Tables 1-4 summarize the results of these analyses for the data of Experiments 1-4, respectively. Values in the tables are averages over the appropriate number of observers in each grouping. Because of difficulty in interpreting the reliability of individual observers' coefficients, conclusions are based on *t* tests on the average coefficients and on the increase in *R*<sup>2</sup> associated with various previous stimuli or responses.

Consider first Tables 1 and 2, which summarize the results from Experiments 1 and 2, in which only judgment method alternated within a single session. These tables reveal the effects of responses made during the performance of one judgment method on responses made during the performance of the other, as well as the effects of current and previous stimuli on current responses.

**Table 1**  
**Regression Coefficients and R<sup>2</sup> Increase from Experiment 1**

Method	Regression Coefficients								
	$\tau$	$\beta_1$	$\alpha_1$	$\beta_2$	$\alpha_2$	$\beta_3$	$\alpha_3$	$\beta_4$	$\alpha_4$
Sound ( <i>n</i> = 4)									
ME	.494†	.126†	-.042†	.121*	-.029	-.013	-.003	.055	-.031
CMM	.356*	.029	-.020	.113*	-.038*	.038*	-.039*	.107	-.031*
Light ( <i>n</i> = 4)									
ME	.301*	.123	-.020	.124*	-.041	-.024	-.007	.145	-.036*
CMM	.249†	.062*	-.030*	.141†	-.036	.042	-.014	.085*	-.025
<b>R<sup>2</sup> Increase</b>									
	<i>I<sub>n</sub></i>	<i>R<sub>n-1</sub></i>	<i>I<sub>n-1</sub></i>	<i>R<sub>n-2</sub></i>	<i>I<sub>n-2</sub></i>	<i>R<sub>n-3</sub></i>	<i>I<sub>n-3</sub></i>	<i>R<sub>n-4</sub></i>	<i>I<sub>n-4</sub></i>
Sound ( <i>n</i> = 4)									
ME	.692	.002	.003	.009	.004	.000	.000	.000	.003
CMM	.706	.002	.006	.010	.016	.000	.003	.004	.006
Light ( <i>n</i> = 4)									
ME	.670	.008	.006	.007	.010	.002	.003	.006	.010
CMM	.586	.002	.011	.026	.022	.002	.003	.004	.009

Note—ME = magnitude estimation. CMM = cross-modality matching. \**p* < .05 by *t* test. †*p* < .01 by *t* test.

**Table 2**  
**Regression Coefficients and  $R^2$  Increase from Experiment 2**

Method	Regression Coefficients								
	$\tau$	$\beta_1$	$\alpha_1$	$\beta_2$	$\alpha_2$	$\beta_3$	$\alpha_3$	$\beta_4$	$\alpha_4$
Sound ( $n = 5$ )									
CJ	.383†	.098	-.043†	.105†	-.053†	-.019	-.024	.074†	-.035†
CMM	.415†	.190*	-.062*	.103*	-.041*	.023	-.033	.070	-.030*
Light ( $n = 3$ )									
CJ	.257*	.095	-.026*	.125*	-.042*	-.008	-.024*	.044	-.016
CMM	.303*	.165	-.045	.166*	-.050*	.066	-.031	.090*	-.029*
$R^2$ Increase									
	$I_n$	$R_{n-1}$	$I_{n-1}$	$R_{n-2}$	$I_{n-2}$	$R_{n-3}$	$I_{n-3}$	$R_{n-4}$	$I_{n-4}$
Sound ( $n = 5$ )									
CJ	.821	.005	.005	.000	.004	.001	.001	.000	.002
CMM	.714	.007	.006	.004	.006	.002	.002	.002	.002
Light ( $n = 3$ )									
CJ	.755	.006	.008	.001	.007	.002	.003	.000	.002
CMM	.652	.004	.014	.021	.022	.000	.003	.005	.005

Note—CJ = category judgment. CMM = cross-modality matching. \* $p < .05$  by  $t$  test. † $p < .01$  by  $t$  test.

For example, in the first row of Table 1,  $R_n$  is a magnitude estimation (ME) of a sound,  $\gamma$  represents the regression coefficient of that sound ( $I_n$ ),  $\beta_1$  is the regression coefficient associated with  $R_{n-1}$ , which was a cross-modality match,  $\alpha_1$  is the regression coefficient associated with  $I_{n-1}$  (also a sound), and so forth. In these tables, even-numbered  $i$  and  $k$  are associated with the same judgment method, and odd-numbered  $i$  and  $k$  are associated with the other judgment method.

Several interesting and informative patterns are revealed in Tables 1 and 2. First,  $\gamma$  represents the usual power function exponent (uncorrected for sequential dependencies). As usual, the exponents for sound pressure amplitude are larger than those for light intensity, although those for sound are somewhat lower than the values considered "typical" (cf. S. S. Stevens, 1975; they are however "typical" of those obtained in my lab). Interestingly, as I found earlier in separate studies (Ward, 1982, 1985, 1986), cross-modality matching exponents are smaller than magnitude estimation exponents but either larger than or roughly equal to category judgment exponents. This is reflected in the relative sizes of the response ranges in the various methods, since a simple power function description of the data such as  $R = aI^\gamma$  implies that  $\gamma = \log R_r / \log R_s$ , where  $R_r$  and  $R_s$  are the response and stimulus range, respectively (see Teghtsoonian, 1971; Ward, 1975). Typical response ranges in magnitude estimation are between 1 and 1.5 log units (about 1.1 log unit in Experiment 1), whereas those for category judgment are usually less than 1 log unit, since seldom are more than 10 categories used (they were 0.9 for sound and 0.8 for light in Experiment 2). Response ranges for cross-modality matching depend on the response continuum; they were between 0.8 and 1.0 in Experiments 1 and 2. Since stimulus ranges were the same for all methods, the

relations between these response ranges roughly correspond to the relations between the exponents. Thus, mixing methods did not appear to make observers use the response continua in an unusual way, nor did it give rise to unusual estimates for the standard power function exponents.

In both tables,  $R^2$  increases for  $I_n$  are large and greater than those for any previous stimulus or response, as they should be for any useful psychophysical method.  $R^2$  increases for magnitude estimation and cross-modality matching are roughly equal, but they are consistently less than those for category judgment, indicating a distinct advantage to the latter method by this criterion. Moreover,  $R^2$  increases for stimuli and responses at lags of 1 or greater are larger for cross-modality matching than for either category judgment or magnitude estimation in these experiments, indicating a drawback of the latter method relative to the other two. This is especially significant, since this is the first time to my knowledge that such measures have been available to be compared across the same observers judging the same stimuli on the same occasion. For purposes of comparison, Ward (1987) found similar relationships among  $R^2$  increases when different groups of observers used the different methods: the average  $R^2$  increases for  $I_n$  were 0.750, 0.650, and 0.480 for category judgment, absolute magnitude estimation, and cross-modality matching, respectively.

Consider next the regression coefficients for lags of 1 or greater. First, all of the  $\alpha_i$  values are negative (6 of 16 significantly so in Table 1, 11 of 16 in Table 2) for both modalities, in both tables, consistent with previous studies that have found negative regression coefficients for previous stimuli in this analysis. One subtlety is that, for cross-modality matching, the  $R^2$  increases associated with stimuli at a lag of 2 are substantially larger than those

at a lag of 1 for three of the four comparisons. Responses at a lag of 2 were made with the same method, while those at a lag of 1 were made with a different method. Variation with response method of the  $R^2$  increase associated with previous stimuli indicates that part of the stimulus contrast effect in cross-modality matching may be of cognitive origin, as argued by Ward (1986).

The regression coefficients for previous responses,  $\beta_k$ , also are generally consistent with previous data. All but 4 of the 32 are positive (8 of 16 are significantly different from zero in Experiment 1, 7 of 16 in Experiment 2; the negative coefficients are all for the effect of cross-modality matches on magnitude estimations or category judgments at a lag of 3, and none are reliably different from 0), indicating reliable assimilation of the current response to previous responses. Interestingly,  $\beta_1$  is substantially larger for magnitude estimation than for cross-modality matching in Experiment 1, but  $\beta_1$  is larger for cross-modality matching than for category judgment in Experiment 2. This indicates that magnitude estimations were assimilated more to immediately previous cross-modality matches than vice versa, while category judgments exerted a larger assimilative effect on subsequent cross-modality matches than vice versa. In Experiment 1,  $\beta_2$  for cross-modality matches is substantially larger than  $\beta_1$ , indicating that a cross-modality match two trials back in the sequence had a larger effect on a current cross-modality match than did a magnitude estimation only one trial back. Nonetheless, the generally (significantly) positive regression coefficients, regardless of lag, indicate that responses in each of the paired methods exerted assimilative effects on responses on the other method, consistent with the fuzzy judgment theory prediction discussed in the introduction.

Tables 3 and 4 display the results of similar analyses for Experiments 3 and 4. One should recall the sequence of trials in these experiments (one complete segment is reproduced as row labels in each table) in order to interpret the regression coefficients and  $R^2$  increases. In general the same conclusions arise, although here the

results are cleaner. Again the exponent for sound pressure was larger than that for light intensity ( $\gamma$ ), and again the exponents for magnitude estimation are the largest for each modality. Relative response ranges also are consistent with these variations of exponents: log response ranges for sound and light were, respectively, 1.5 and 1.3 for magnitude estimation, 0.9 and 0.8 (Experiment 3) and 0.7 and 0.7 (Experiment 4) for cross-modality matching, and 0.8 and 0.8 for category judgment. Again, category judgment is superior to both other methods by the criterion of larger  $R^2$  increase associated with  $I_n$ , and smaller  $R^2$  increase associated with stimuli at lags of 1 or more, although magnitude estimation also fares well in Experiment 3. Cross-modality matching (at least using the duration response continuum) is the clear loser in both experiments.

In the case of the  $\alpha_i$  coefficients, we can examine both intramodal and intermodal effects, since modalities alternated in these experiments. Intramodal effects arise from previous trials on which the same modality was judged (as that on the current trial), whereas intermodal effects arise from previous trials on which the other modality was judged. For example, in row ME-S<sub>1</sub> of Table 3,  $\alpha_1$  represents the effect of the stimulus on the immediately preceding trial, which was a CMM-L<sub>2</sub> trial, on an ME-S<sub>1</sub> judgment, an intermodality effect (light-on-sound). In the same row,  $\alpha_2$  represents the effect of ME-L<sub>1</sub> trials, also an intermodality effect, while  $\alpha_3$  and  $\alpha_4$  represent the effects of CMM-S<sub>2</sub> and ME-S<sub>1</sub> trials, respectively, both intramodality effects, on the same ME-S<sub>1</sub> judgment. There are 16 intramodal effects of previous stimuli displayed in the two tables, and all are negative and significantly different from zero. The intermodality coefficients are small; overall, 10 of 16 are negative. Only 2 of them (both negative) are significantly different from zero; they are on same-method trials for cross-modality matches in Experiment 3 ( $\alpha_2$  for CMM-S<sub>2</sub> and CMM-L<sub>2</sub> trials), consistent with Ward's (1986) finding of a small intermodality contrast effect for cross-modality matches. Overall, the data support the previous finding that stimulus con-

**Table 3**  
Regression Coefficients and  $R^2$  Increase from Experiment 3

Method and Stimulus Modality	Regression Coefficients								
	$\tau$	$\beta_1$	$\alpha_1$	$\beta_2$	$\alpha_2$	$\beta_3$	$\alpha_3$	$\beta_4$	$\alpha_4$
ME-S <sub>1</sub>	.639†	-.001	.016	.054	-.005	.085	-.082†	.093	-.080*
CMM-S <sub>2</sub>	.351†	.148†	-.111†	.204†	-.036†	.010	-.001	.191†	-.093†
ME-L <sub>1</sub>	.451†	.053	-.014	.068†	-.033	.084†	-.057†	.106†	-.059†
CMM-L <sub>2</sub>	.285†	.052*	-.039*	.177†	-.048†	-.026	.011	.182†	-.058†

	$R^2$ Increase								
	$I_n$	$R_{n-1}$	$I_{n-1}$	$R_{n-2}$	$I_{n-2}$	$R_{n-3}$	$I_{n-3}$	$R_{n-4}$	$I_{n-4}$
ME-S <sub>1</sub>	.737	.002	.002	.005	.002	.002	.007	.001	.003
CMM-S <sub>2</sub>	.636	.005	.027	.027	.011	.000	.001	.003	.018
ME-L <sub>1</sub>	.735	.002	.002	.002	.001	.003	.005	.002	.004
CMM-L <sub>2</sub>	.635	.002	.008	.016	.016	.001	.001	.003	.007

Note— $n = 8$ . ME = magnitude estimation. CMM = cross-modality matching. S = sound. L = light. \* $p < .05$  by  $t$  test. † $p < .01$  by  $t$  test.



**Table 4**  
**Regression Coefficients and R<sup>2</sup> Increase from Experiment 4**

Method and Stimulus Modality	Regression Coefficients								
	$\tau$	$\beta_1$	$\alpha_1$	$\beta_2$	$\alpha_2$	$\beta_3$	$\alpha_3$	$\beta_4$	$\alpha_4$
CJ-S <sub>1</sub>	.365†	.072*	-.009	.055	-.006	.024	-.039†	.047	-.041*
CMM-S <sub>2</sub>	.297†	.242†	-.080†	.125*	-.018	.034	-.005	.110	-.061*
CJ-L <sub>1</sub>	.271†	.059	-.003	.056	-.004	.071†	-.040†	.077	-.028†
CMM-L <sub>2</sub>	.228†	.207†	-.045*	.177†	.069	-.013	.006	.134†	-.049*

	R <sup>2</sup> Increase								
	$I_n$	$R_{n-1}$	$I_{n-1}$	$R_{n-2}$	$I_{n-2}$	$R_{n-3}$	$I_{n-3}$	$R_{n-4}$	$I_{n-4}$
CJ-S <sub>1</sub>	.802	.001	.000	.001	.001	.005	.003	.002	.002
CMM-S <sub>2</sub>	.701	.007	.014	.007	.005	.002	.001	.002	.010
CJ-L <sub>1</sub>	.786	.004	.001	.003	.001	.003	.008	.002	.002
CMM-L <sub>2</sub>	.646	.020	.011	.013	.006	.001	.002	.001	.011

Note— $n = 6$ . CJ = category judgment. CMM = cross-modality matching. S = sound. L = light. \* $p < .05$  by  $t$  test. † $p < .01$  by  $t$  test.

trast effects are largely modality-specific in mixed-modality scaling situations. The mixed-method context does not alter this pattern.

Again, similarly to previous findings, assimilative sequential dependencies do not depend heavily on modality, although as in Experiments 1 and 2, there are some differences in the effects of responses in the various methods on one another. Twenty-nine of 32  $\beta_k$  coefficients are positive, and 16 of these are significantly different from zero. The 3 negative coefficients are small, not significantly different from zero, and with one exception, they occur at higher lags. The single exception is interesting. It is  $\beta_1 = -0.001$ , for the ME-S<sub>1</sub> judgment (Table 3). It indicates essentially no effect of the immediately previous CMM-L<sub>2</sub> response on that judgment. In general, in Experiment 3, cross-modality matches had less assimilative effect on magnitude estimations than vice versa, and in Experiment 4, the same held for cross-modality matches and category judgments. The influence of previous category judgments on current cross-modality matches was even greater than the influence of previous cross-modality matches in Experiment 4, whereas previous cross-modality matches had more influence on current ones in Experiment 3, even though they occurred at a greater lag. In general, these results are also consistent with the idea (discussed in the introduction) of a common categorical basis of judgment in the three methods.

**Second-order dependencies.** Figures 5 and 6 illustrate the results of analyses of the data of Experiments 1 and 2 for the dependency of the correlation between previous and present responses on the separation between the stimuli responded to ( $S_n - S_{n-k}$ ). This dependency typically has the form of an inverted V centered at  $S_n - S_{n-k} = 0$  (see Ward, 1982, 1986, for data from comparable methods). As can be seen in the figures, this pattern is present for both intramethod and intermethod assimilative dependencies (positive correlations). This means that the assimilative effects of responses in one method on those in another have the same form as those of each method on themselves. The results of such analyses for Experiments

3 and 4 are similar but more variable (because of the smaller numbers of trials on which individual correlation coefficients are based) and are not shown. Again, this result is consistent with a common source for the assimilative dependencies in the process of categorization assumed by the fuzzy judgment theory to be common to all three methods.

### CONCLUSION

The results of the experiments reported above support a number of useful conclusions. First, it appears that mixing methods of judgment does not significantly distort how either method is used. This is true of ranges of responses obtained, exponents of power functions fitted to the data, derived cross-modality matching functions, and dependencies of current responses on previous stimuli and responses of the same method and the same or different modality. In the mixed-method context, as in others (e.g., Ward, 1987), category judgments provide more consistent judgments, probably because of the strong constraints that the method imposes on the responses that can be used. Finally, derived cross-modality matching functions are somewhat cleaner in the mixed-method context but have comparable slopes to those derived in the simple mixed-modality context. Thus, mixed-method scaling, especially when combined with mixed-modality scaling, has been shown to be at least as valid as other direct scaling methods. In addition, it is considerably more efficient, in that it produces equally useful psychophysical functions while at the same time producing additional data, such as response matching functions, that give insight into the judgment process and provide a basis for assessment of internal consistency.

The present experiments have also provided additional evidence that magnitude estimation, category judgment, and cross-modality matching have much in common. Apparently responses are used in the various methods in a way that allows relatively simple equivalences to be established between them, although they are never simply

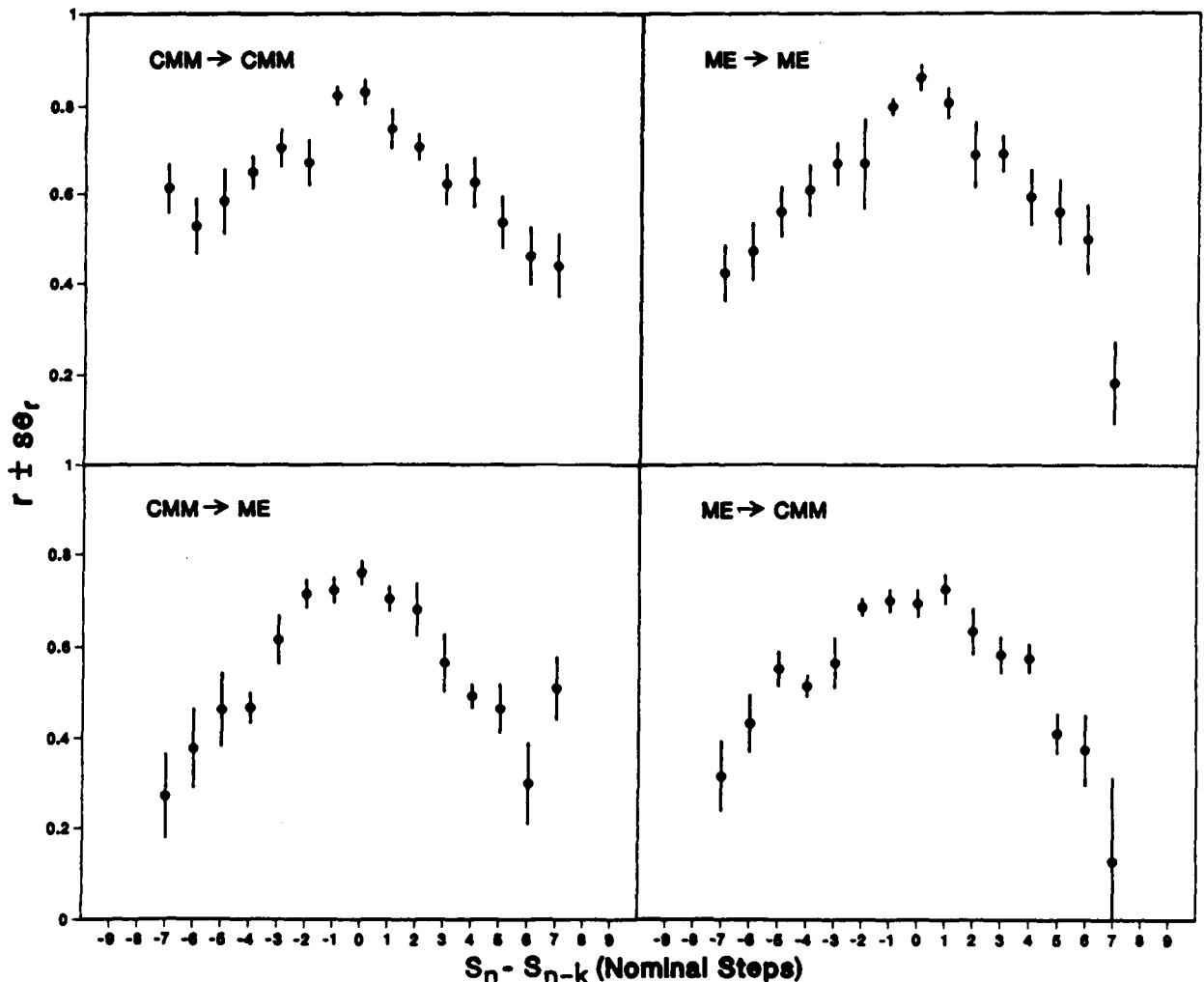


Figure 5. Plots of the correlation between  $\log R_n$  and  $\log R_{n-k}$  as a function of  $S_n - S_{n-k}$  for lags of 1 ( $k = 1$ ) and 2 ( $k = 2$ ) for the data of Experiment 3.

translations of number to a cross-modality matching response or vice versa. While maintaining different ranges of responses with the different methods, observers nevertheless use these responses consistently enough for simple power or linear functions to describe their relationship adequately. Moreover, dependencies of current responses on previous responses usually have the same form regardless of whether the method of judgment producing the previous response was the same as or different from that producing the current one. This does not seem to arise from a simple biasing of one method by the other (stronger?) method, since each method retained its characteristic properties in the present experiments.

All of this is consistent with the idea that responses in the three methods share a common stage of judgment. A plausible common mechanism is categorization, as described in the theory of fuzzy judgment. This theory explains the patterns of sequential dependencies observed in the present experiments by attributing intramodal con-

trastive dependencies on previous stimuli to a stimulus discrimination process while assimilative dependencies on previous responses are said to arise from the use of heuristics applied during the categorization process. The fact that the methods retained their usual characteristics vis-à-vis power function exponents and response ranges is explained by the fact that the output from the categorization process becomes the input to a response process that differs with method, depending on the rule of judgment imposed by the experimenter.

Interestingly, S. S. Stevens (1975) argued that cross-modality matching could not be mediated by magnitude estimations or any other numerical judgment (such as judgment based on a physical scale of the stimuli, for example), because cross-modality matching seemed so simple and direct *experientially*, and because naive observers, including young children who had never used magnitude estimation and were numerically unsophisticated, could immediately and directly match sensation intensity on two

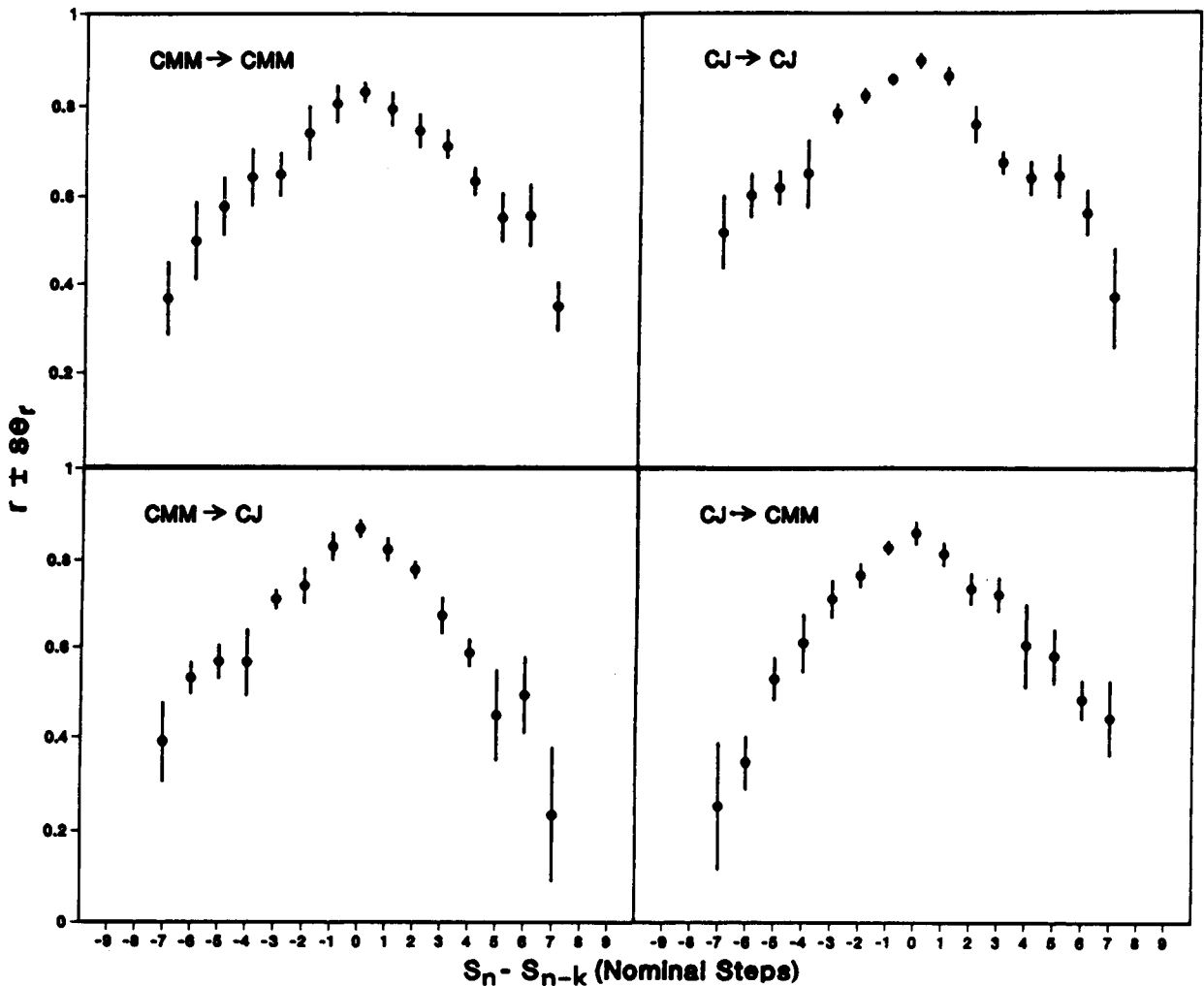


Figure 6. Plots of the correlation between  $\log R_n$  and  $\log R_{n-k}$  as a function of  $S_n - S_{n-k}$  for lags of 1 ( $k = 1$ ) and 2 ( $k = 2$ ) for the data of Experiment 4.

continua. However, categorization can be nonnumerical; it is a primitive operation performed by all of us on a regular basis; and it often (if not always) provides the basis of perceptual equivalence that is required for effective action. Thus, it is possible that mediation of cross-modality (and intramodality) matches by a process of categorization can explain Stevens's informal observations of the ease and simplicity of cross-modality matching, as well as the many properties of cross-modality matching that he discovered (S. S. Stevens, 1975, chap. 4).

The final picture is not completely clear. There are asymmetries and inconsistencies in the effects of judgment methods on each other. For example, magnitude estimations were more affected by immediately previous cross-modality matches than vice versa in Experiment 1, but affected only weakly by them in Experiment 3 (not at all for the ME- $S_1$  judgments). The different groups of observers in Experiments 1 and 2 seem to have used their responses somewhat differently. Clearly, both the ex-

perimental context and decisions made by different groups of observers about how to comply with the experimenter's instructions could have significant effects on the outcomes of such experiments. The challenge is to classify the various effects as to source correctly, so that a valid, fundamental, and method-accommodating theory of psychophysical judgment can emerge.

REFERENCES

AFTANAS, M. S., & RULE, S. J. (1968). Effect of cross-continuum anchors. *American Journal of Psychology*, *81*, 559-565.  
 JESTEADT, W., LUCE, R. D., & GREEN, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology: Human Perception & Performance*, *3*, 92-104.  
 KRUEGER, L. E. (1989). Reconciling Fechner and Stevens: Toward a unified psychophysical law. *Behavioral & Brain Sciences*, *12*, 251-267.  
 LEVINE, D. S., & GROSSBERG, S. (1976). Visual illusions in neural networks: Line neutralization, tilt aftereffects, and angle expansions. *Journal of Theoretical Biology*, *61*, 477-504.  
 LOCKHEAD, G. R., & KING, M. C. (1983). A memory model of se-

- quential effects in scaling tasks. *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 461-473.
- LUCE, R. D., BAIRD, J., GREEN, D. M., & SMITH, A. F. (1980). Two classes of models for magnitude estimation. *Journal of Mathematical Psychology*, **22**, 121-148.
- MARKS, L. E., SZCZESUIL, R., & OHLLOTT, P. (1986). On the cross-modal perception of intensity. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 517-534.
- MARLEY, A. A. J., & COOK, V. (1986). A limited capacity rehearsal model for psychophysical judgments applied to magnitude estimation. *Journal of Mathematical Psychology*, **30**, 339-390.
- PARKER, S., CASEY, J., ZIRIAX, J. M., & SILBERBERG, A. (1988). Random monotone data fit simple algebraic models: Correlation is not confirmation. *Psychological Bulletin*, **104**, 417-423.
- STEVENS, J. C., & MARKS, L. E. (1980). Cross-modality matching functions generated by magnitude estimation. *Perception & Psychophysics*, **27**, 379-389.
- STEVENS, S. S. (1975). *Psychophysics*. New York: Wiley.
- TEGHTSOONIAN, R. (1971). On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review*, **78**, 71-80.
- TREISMAN, M., & WILLIAMS, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, **91**, 68-111.
- WARD, L. M. (1975). Sequential dependencies and response range in cross-modality matches of duration to loudness. *Perception & Psychophysics*, **18**, 217-223.
- WARD, L. M. (1979). Stimulus information and sequential dependencies in magnitude estimation and cross-modality matching. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 444-459.
- WARD, L. M. (1982). Mixed-modality psychophysical scaling: Sequential dependencies and other properties. *Perception & Psychophysics*, **31**, 53-62.
- WARD, L. M. (1985). Mixed-modality psychophysical scaling: Inter- and intramodality sequential dependencies as a function of lag. *Perception & Psychophysics*, **38**, 512-522.
- WARD, L. M. (1986). Mixed-modality psychophysical scaling: Double cross-modality matching for "difficult" continua. *Perception & Psychophysics*, **39**, 407-417.
- WARD, L. M. (1987). Remembrance of sounds past: Memory and psychophysical scaling. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 216-227.
- ZADEH, L. (1965). Fuzzy sets. *Information & Control*, **8**, 338-353.

## NOTE

1. In this paper, the traditional method of fitting power and linear functions to scaling data and assessing such fit is used. This method relies heavily on linear regression and the  $r^2$  measure of goodness of fit. In a recent paper, Parker, Casey, Ziriak, and Silberberg (1988) demonstrated that random monotone data fit linear and power functions with quite high  $r^2$  values (95th percentiles above 0.95). Moreover, since their data were generated with an arbitrary selection of sampling scheme and ordinate geometry, even if some experimental data produce  $r^2$  values above the 95th percentile (which all of the present data do), it is possible that some other arbitrary set of assumptions would yield equivalent fits for random data. Thus, Parker et al. (1988) concluded that  $r^2$  alone is inadequate as a test of fit of monotone data to algebraic models. In light of these rather iconoclastic conclusions, it should be emphasized that in the present paper, linear and power function models are used only *descriptively*. That is, such algebraic models are fitted to the data in order to describe them economically and to compare them with previous data described with recourse to the same empirical models. There is no intention of asserting that such models are superior to other algebraic models for theoretical or empirical purposes on the basis of the reported  $r^2$  measures. It is possible that other models would prove superior, particularly if they explained some of the (not dramatic) nonlinearities. It is also true that merely to use such models without comparing them to others tends to enhance their reputation for usefulness at the expense of the (untested) others. However, linear and power function models do provide a good first approximation to most psychophysical data (including those presented in this paper), and they are simple enough to work with easily. For this reason, they have been used in the present paper, in which the intent is to describe and begin to interpret some new data.

(Manuscript received May 25, 1989;  
revision accepted for publication June 25, 1990.)