

Memory for psychophysical scaling judgments

LAWRENCE M. WARD

University of British Columbia, Vancouver, British Columbia, Canada

JULIET ARMSTRONG

Simon Fraser University, Burnaby, British Columbia, Canada

and

NARLY GOLESTANI

McGill University, Montreal, Quebec, Canada

In three experiments, we examined memory for responses and stimuli experienced in a single direct psychophysical scaling session in which subjects made absolute magnitude estimations of the loudnesses of pure tones. Recall of scaling responses was found to be accurate for the softest and loudest stimuli, but systematically greater than actual judgments for the intermediate stimulus amplitudes, yielding distorted psychophysical functions for the recall data which nonetheless had the same power function exponent as that for the judged stimuli. Also, memory for the range of stimulus amplitudes was fairly accurate, but subjects could not distinguish between judged and nonjudged amplitudes within that range. The results are consistent with the role of extreme stimuli as anchors for judgment, and indicate that memories for these stimuli and responses made to them can be expected to influence future scaling judgments. These results also are consistent with the uncertainty hypothesis of mnemophysics.

Memory and perception have been intertwined since Plato and Aristotle wrote, and Fechner applied his psychophysical approach not only to perceptual continua but also to higher mental processes such as memory (see Algom, 1992, for a review). Helson (e.g., 1964) proposed that the effects of previously encountered stimuli, both outside of and within a psychophysical scaling session, on the adaptation level would strongly affect the way a current stimulus was perceived and judged.

Ward (1987) summarized these and other memory effects in psychophysical scaling (e.g., Marks, Szczesiul, & Ohlott, 1986; M. Teghtsoonian & R. Teghtsoonian, 1983; Ward, 1979) by proposing that experimental subjects establish several different *fuzzy maps* between stimuli and responses. He found that memories of stimulus sets experienced on previous days strongly influenced psychophysical scaling judgments of current stimuli and that current judgments seemed to be based on a compromise between relative and absolute maps formed at least partly in previous scaling sessions. However, it was unclear from these studies just what was remembered from the previous scaling sessions and just how memories of previous stimuli and responses gave rise to biased scaling judgments.

Memory and psychophysical scaling are also closely linked in what is called memory psychophysics or *mnemo-*

physics (Algom, 1992). In mnemophysics, subjects are asked to make psychophysical scaling judgments of *memories* of previously experienced stimuli. At least two important findings have emerged from studies of mnemophysics. First, psychological magnitudes of remembered stimuli are related to the physical magnitudes of those stimuli in the same way as psychological magnitudes of currently perceived stimuli are: by power functions. Second, the exponents of these power functions are usually lower for remembered magnitudes than they are for currently perceived magnitudes. For example, the power function exponent for remembered area in visual space is about 0.46, whereas that for perceived area is about 0.64 (Moyer, Bradley, Sorensen, Whiting, & Mansfield, 1978; see Algom, 1992, for exceptions).

There are two major theoretical explanations of the typical findings. The *reperception hypothesis* states that in the judging of remembered stimulus magnitudes, the memories of the stimuli are "reperceived" using the same power function transform as when the stimuli were originally perceived, and thus applying the transform twice (Kerst & Howard, 1978; Moyer et al., 1978). This leads to a squaring of the exponent for judgments of memories. For perceptual exponents smaller than 1, this results in a decrease in exponent magnitude (0.64 squared equals 0.41), whereas for those larger than 1, it results in a larger exponent (1.5 squared equals 2.25; cf. Algom & Lubel, 1994). A hybrid reperception model suggests that the ratio of memory and perceptual exponents is the "true" perceptual exponent (Algom & Marks, 1989).

The *uncertainty hypothesis* holds that exponents are smaller in the memory task because uncertainty about re-

The present research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to L.M.W. Thanks to the reviewers for insightful critiques. Correspondence should be addressed to L. M. Ward, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC V6T 1Z4, Canada (e-mail: lward@cortex.psych.ubc.ca).

membered stimulus magnitudes is greater than uncertainty about perceived ones. This could lead either to response range constriction (Kerst & Howard, 1978) or to changes on the sensory continuum that are reflected in the responses (Algom, Wolf, & Bergman, 1985). Either effect implies smaller memory exponents, whether the perceptual exponent is less than 1 or greater than 1.

Unidimensional mnemophysical studies have not resolved all the problems that arise in the context of these theories and results (Algom, 1992). A persisting problem, at least for the unidimensional studies, is that the usual assumption about the perceptual scaling context, that it is "pure" perception, is not correct: The perceptual exponents are also affected by memories of previous judgments and stimuli, as has been described above. Thus, in order to correctly interpret the mnemophysical results, we need to know more about the influence of memory on the basic psychophysical results.

It is also important to know what subjects remember from psychophysical scaling situations because tests of consistency of exponents across time and situations make assumptions about memory for the stimuli encountered, the responses made, and the relation between them developed in previous scaling sessions (Marks, 1991; M. Teghtsoonian & R. Teghtsoonian, 1983; Verillo, 1983). For example, Marks (1991) pointed out that if scaling subjects remember only the highest and lowest responses given but not the associated perceptual magnitudes, then both high intra-individual reliability of responses to the same stimuli in different sessions and large effects of context on responses to different stimuli would be expected. Unfortunately, no direct studies of memory for psychophysical scaling judgments exist to confirm this conjecture.

Finally, models of psychophysical judgment often contain a memory component. For example, J. A. Siegel and W. Siegel (1972) argued that absolute judgment of sensory magnitudes is like paired-associate learning, in which performance is dominated by remembered relations between stimuli and responses. This, however, seems implausible if perceptions of sensory stimuli are represented in memory not as precise quantities but rather as fuzzy subsets (Ward, 1979). Paired-associate learning is an explicit, episodic form of memory or, at the very least, a form of explicit memory for gist. Stimulus-response maps formed on this basis would also be precise and explicit. Neisser's (1981) characterization of *repisodic* memory seems to fit the psychophysical scaling situation better: Repisodic memory is memory for the commonalities (in this case, fuzzy S-R mapping) of repeated episodes without any precise memory of specific episodes (Ward, 1987).

It is possible, however, that subjects might remember a few particular responses or experienced sensory magnitudes, since we know that in mnemophysical studies subjects give reliably different responses to different remembered sensory experiences. But *which* responses and stimuli do they remember? And how well do they remember them? Also, from the literature, we might expect

that the likelihood of distinguishing in memory between two stimuli should increase as the similarity between them decreases. We know nothing, however, about how dissimilar such stimuli would have to be to be distinguished in memory. Stimuli in scaling studies are inherently highly similar—usually differing only on a single physical dimension that leads to differences in sensory intensity. Would it suffice that they be perceptually distinguishable? Or would they need to be separated by 5, 10, 20, or even 100 times the difference threshold? The answers to such questions are unknown. Thus, guided by the questions above, the present studies were designed to provide the first direct measurements of what subjects remember of psychophysical scaling situations.

EXPERIMENT 1

In Experiment 1, we asked our subjects to give responses in a typical psychophysical scaling situation, absolute magnitude estimation of loudness of pure tones, and then asked them to recall the responses they had given to each of the pure tones during that session, without re-presenting the tones. We tested recall under one of two sets, *incidental* or *intentional*, and after a delay of either 5 min (5m) or 1 day (1d). Since we focused on memory for S-R mappings, we did not test memory for stimuli directly in this experiment. However, as will become clear, we presumed that some representation of the stimuli also would be involved in the responses that subjects gave in the recall task.

Method

Subjects. Sixteen volunteer University of British Columbia undergraduate students participated for pay. All claimed to have normal hearing and had no difficulties hearing the tones used as stimuli. Seven were females and 9 were males.

Stimuli and Apparatus. All stimuli were 1000-Hz, 1-sec duration sinusoids, generated, amplified, and electronically gated (2.5-msec rise and fall times) by a custom-built digitally controlled sound generator, and controlled and timed by a microcomputer that also recorded the subjects' responses. The tones were presented monaurally through Koss Pro-4AAA earphones while the subjects sat in a sound attenuation chamber. Responses were made on a standard computer keyboard. The target stimuli consisted of eight tones with amplitudes of 40–96 dB in 8-dB steps. All the stimuli were measured at the earphone with a custom-built artificial ear and a General Radio precision sound-level meter.

Procedure. Each subject first participated in a 200-trial psychophysical scaling session, preceded by a brief (approximately 20-trial) practice run, with stimulus amplitudes presented in a pseudo-random sequence. For this scaling session, the subjects were required to give to each stimulus a numerical response whose apparent magnitude matched that of the stimulus (cf. Zwislocki & Goodman, 1980). The subjects in the intentional groups (Intentional-5m and Intentional-1d, $n = 4$ in each group) were also told that they would be asked to recall all of the responses that they had given to each stimulus; the subjects in incidental groups (Incidental-5m and Incidental-1d, $n = 4$ in each group) were told nothing further. Approximately 5 min or 24 h after completion of the scaling session, the subjects were asked to recall all of the 200 numerical judgments that they had given in that session. They were told that there had been eight different tone amplitudes, each of which they had heard

25 times in a random order, that they did not always give the same response to the same amplitude, and that they gave the same response to more than one amplitude. They were required to fill in all 200 cells on the answer sheet, which had a row for each of the eight tones and a column for each of the 25 presentations of each tone. The order of recall was left unspecified and was not constrained to be the same as the order of stimulus presentation in the scaling session.

Results and Discussion

A between-within analysis of variance (ANOVA: between, recall set and delay; within, stimulus amplitude and judged vs. recalled) was performed on the logarithms of subjects' geometric mean responses (Huynh-Feldt adjusted df for amplitude). Log geometric mean responses were analyzed because psychophysical scaling response distributions tend to be highly skewed, and also because linear regressions are typically performed on such transformed responses in order to estimate parameters of psychophysical functions (see later). There was no significant effect of recall set [$F(1,12) < 1.0$] or delay [$F(1,12) < 1.0$], and there were no interactions involving these factors. However, judged versus recalled [$F(1,12) = 14.13, p < .003$] and stimulus amplitude [$F(7,84) = 63.83, p < .0001$] and their interaction [$F(7,84) = 7.13, p = .0001$] were significant.

Average judged and recalled psychophysical functions, collapsed across conditions and individuals, are plotted on log-log axes in Figure 1. The judged function forms a very nice straight line, whereas the recalled function is

distinctly curved, consistent with the significant interaction reported earlier. The curvature of the recalled function arises because subjects accurately recalled average responses to the two extreme stimuli whereas they consistently overestimated average responses to the intermediate amplitudes.

We also fitted functions of the form (log-transformed power function)

$$\log R = m \log S + \log a \quad (1)$$

to both the judged and the recalled psychophysical functions of individual subjects, using the log geometric mean responses and log stimulus amplitudes (S in dynes/cm²). Note that because of the departure of the recalled function from a straight line in the log-log plot (Figure 1), the fit to these data is not optimal and has been done only for purposes of comparison with usual scaling results. An ANOVA of the resulting estimates of m , which represents the exponent of the psychophysical power function for the nontransformed responses and stimuli, revealed no significant difference between subjects' judged and recalled functions [$F(1,12) < 1.00; m_j = 0.33$ and $m_r = 0.32$] and no significant differences or interactions across delay or recall set. Intercepts ($\log a$, where a represents the multiplicative constant of the power function) were significantly higher for recalled functions than for judged functions [$F(1,12) = 20.46, p < .001$] but did not differ significantly across recall set or delay. Regression

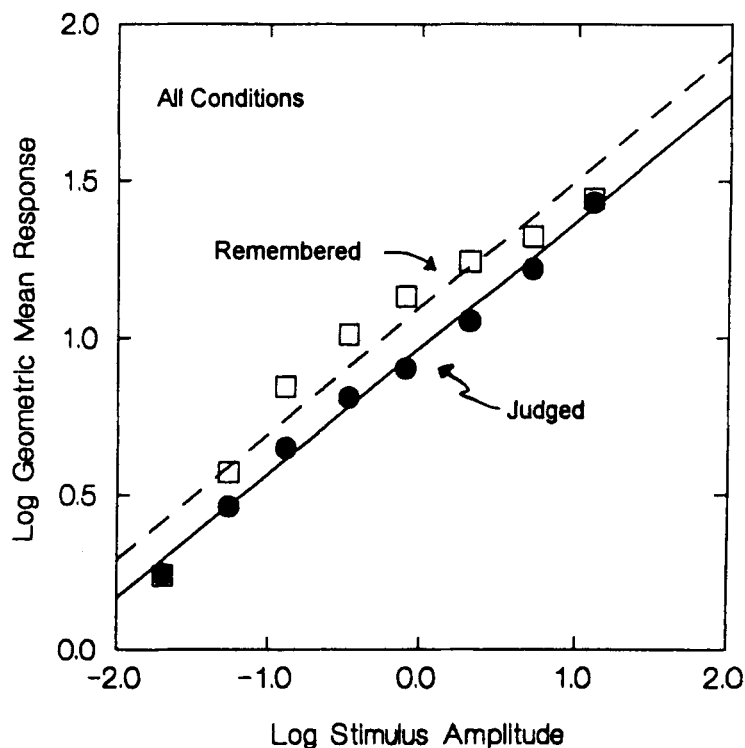


Figure 1. Average psychophysical functions constructed from recalled and judged responses from Experiment 1. The straight lines are the best-fitting lines described in the text.

analyses of the average functions in Figure 1 gave similar results: For the judged function, adjusted $r^2 = .992$, $m_j = 0.402$, and $\log a = 0.971$, while for the recalled function, adjusted $r^2 = .927$, $m_r = 0.406$, $\log a = 1.101$. Interestingly, Algom (1992) reported three studies that yielded similar results for loudness, both judged and memory exponents of approximately 0.4. However, both the present results and those reported by Algom differ from most results in memory psychophysics. Perhaps loudness, an intensive continua, is remembered differently from the extensive continua, such as line length, that are often the subjects of mnemophysical studies.

That no main effect was found for recall set (incidental vs. intentional instructions), or delay (5 min vs. 1 day), is suggestive of episodic memory (see the introduction), but not of episodic memory of specific associations between stimuli and responses, in which measurable decay would be expected over a 24-h period. Further support for this notion is provided by the fact that the subjects were able to reproduce the slope of their psychophysical functions, but not the mean responses that they gave to intermediate stimuli. For the observed performance, including the curvature of the psychophysical function in log-log coordinates, it would suffice that subjects roughly recalled their responses to the extreme stimuli and divided the response range between the extremes into the required number of equal increments. This would make average responses increase linearly with log stimulus amplitude for the intermediate stimuli, since stimuli were log-equally spaced. This is consistent with the fact that the recalled responses form a better straight line in semilog coordinates (geometric mean responses vs. log stimulus amplitudes: adjusted $r^2 = .975$) than in log-log coordinates (adjusted $r^2 = .927$). The subjects did not remember the exact correspondence between stimulus amplitudes and their judgments for the intermediate stimuli, possibly because they did not remember exactly what the intermediate stimuli sounded like. However, it is possible that the subjects did not remember what the extreme stimuli sounded like either, but only remembered their extreme responses, since that would have sufficed to produce the result of Experiment 1. In order to differentiate between these two possibilities, we ran Experiments 2 and 3.

EXPERIMENTS 2 AND 3

We tested incidental recognition memory for particular stimulus amplitudes experienced in a scaling session in two ways. First, we asked subjects to respond "yes" or "no" as to whether a particular sound was one of the sounds that they had judged in a previous scaling session (*recognition task*). Then, each subject also made a series of similar "yes"/"no" judgments on a set of sounds presented repeatedly, so that an unbiased index of recognition memory for the previously judged stimulus amplitudes (d' from signal detection theory; see Green & Swets, 1974) could be calculated (*discrimination task*). If subjects' recall performance in Experiment 1 was based only

on memories for the extreme responses and all else was inference, they might remember none of the stimulus amplitudes that they had judged. If it was based on a relationship between remembered sensation and response amplitudes for the extreme stimuli only, subjects should know at least whether a given sound intensity was inside or outside the range of amplitudes used in the scaling session. Finally, if it was based on a remembered relationship between particular sensations and particular responses across the entire range of stimulus amplitudes, even if the remembered relationship was distorted, then subjects might be able to discriminate among sound amplitudes within that range as to whether or not they had been judged in the earlier session.

Method

Subjects. Eight volunteer University of British Columbia undergraduate students in Experiment 2 (6 female and 2 male) and 10 in Experiment 3 (3 male and 7 female) participated for pay. All claimed to have normal hearing, and none in Experiment 2 had any difficulties hearing the tones used as stimuli. The data of 4 additional subjects in Experiment 3 were discarded without analysis, because they had difficulty hearing the lowest amplitude tones.

Stimuli and Apparatus. The apparatus was identical to that used in Experiment 1. In Experiment 2, the recognition task stimulus set consisted of the following tone amplitudes (scaling, or target, tones are given in bold): 20, 25, 30, 35, **40**, 42.5, **45**, 47.5, **50**, 52.5, **55**, 57.5, **60**, 62.5, **65**, 67.5, **70**, 72.5, **75**, 80, 85, 90, and 95 dB. The 20-, 25-, 30-, 85-, 90-, and 95-dB tones were eliminated for the discrimination task. In Experiment 3, the recognition and discrimination task stimulus sets were (scaling targets in bold): 35, **40**, **45**, **50**, 55, **60**, 65, **70**, 75, **80**, 85, **90**, and 95 dB.

Procedure. Each subject first participated in a 200-trial (210 in Experiment 3) scaling session identical to that for Experiment 1. For Experiment 2, 4 subjects did the 23-trial recognition session 5 min after the scaling session concluded and then immediately did the 160-trial discrimination session; the other 4 did the recognition and discrimination sessions approximately 24 h after the scaling session. For Experiment 3, all subjects performed both recognition and discrimination tasks after a 5-min delay.

In Experiment 2, each of the 23 recognition set stimuli was judged once in a different pseudo-random sequence for each subject. The subjects were told to try to identify the tones that they had judged in the scaling session, and they were told that some of the tones that they would hear would be those tones, and that some would be tones that they had not judged. In the discrimination task, the subjects judged each of the 17 discrimination set stimuli about 9–10 times in a different pseudo-random order for each subject. They were told that the task was identical to the recognition task but that they would be presented with targets and distractors several times each.

In Experiment 3, the subjects were told in addition that there would be distractor tones near each of the target tones, including those inside the range of target tone amplitudes. In the discrimination phase, the 35- and 95-dB distractors were each judged 7–8 times and all other tones were judged 15 times each, equalizing the number of judgments per tone for inside-range and outside-range stimuli.

Results and Discussion

For both Experiments 2 and 3, the subjects' scaling data were analyzed as in Experiment 1 above. They were normal in all respects and approximately the same as those reported for Experiment 1. For brevity, they are not reported here.

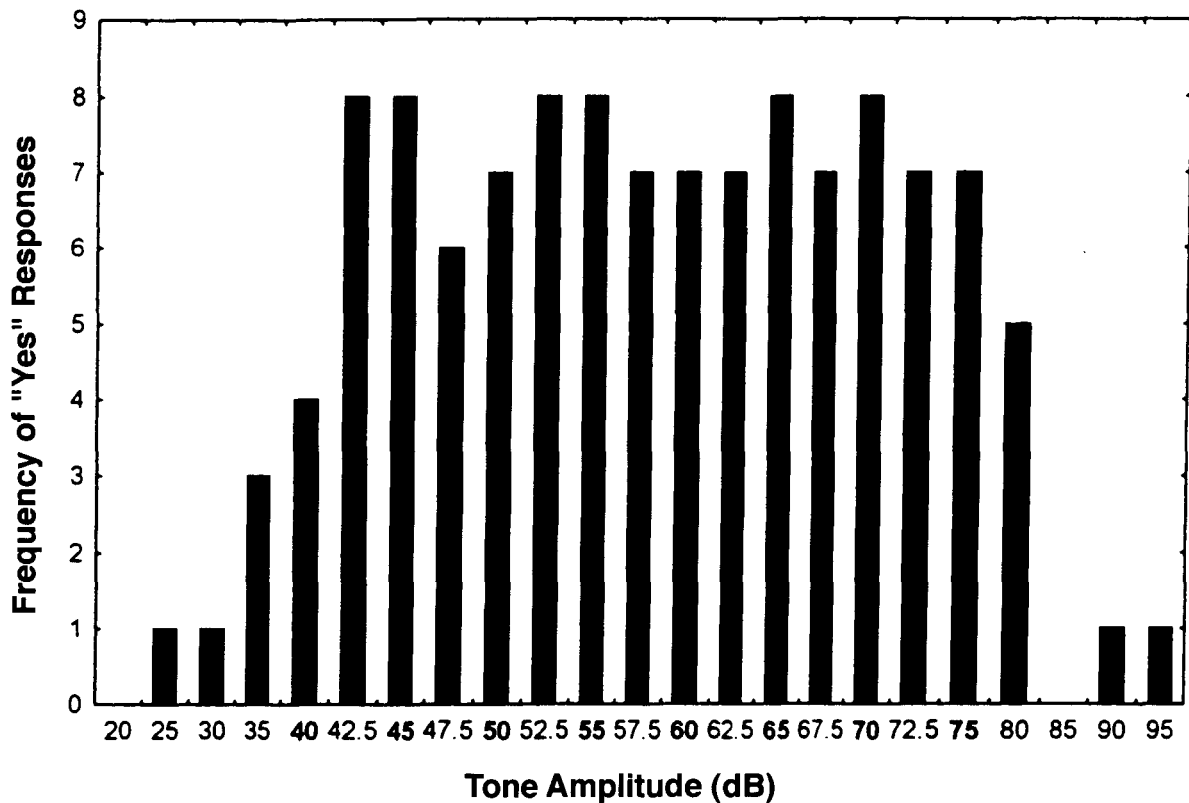


Figure 2. Numbers of subjects in Experiment 2 who said "Yes, that tone was one of the tones I judged in the previous session" for each of the 23 stimuli in the recognition set. Target tone amplitudes are in boldface.

Figure 2 displays the total frequency (over the 8 subjects) of affirmative responses to each of the 23 stimuli in the recognition set for Experiment 2. Figure 1 suggests that subjects were able to ascertain roughly whether a particular sound amplitude was or was not inside the range of the target amplitudes, but were not able to discriminate among amplitudes within this range as to whether they had or had not been judged previously.

A signal detection theory analysis was conducted on the data from the discrimination task in Experiment 2. The index of discriminability, d' , was calculated on the basis of the numbers of hits and false alarms for all subjects combined, since the d' values for individual subjects were not stable with so few responses. A *hit* was defined as a "yes" given to a target stimulus and a *false alarm* as a "yes" given to a distractor. There were no reliable differences between the delay conditions, so that factor is ignored in what follows. The overall hit rate was used with the within-range false alarm rate (excluding 35-dB and 80-dB distractors) to calculate d'_1 , and with the overall false alarm rate (all distractors) to calculate d'_2 .

Based on all subjects' responses, $d'_1 = -0.30$ and $d'_2 = -0.05$. The d'_1 value is significantly different from zero ($z = 3.95$, $p < .001$; see Macmillan & Creelman, 1991) and from the d'_2 value, but it is negative. The d'_2 value is not different from zero. The false alarm rate for outside-range distractors (.12 and .42, respectively, for 35- and

95-dB distractors) is significantly lower than that for inside-range distractors (.82 overall), lowering the overall false alarm rate to .76 (the hit rate was .73) for d'_2 . The lower outside-range false alarm rate again indicates that the subjects were only identifying outside-range distractors as different from judged stimuli. Apparently, the subjects were not able to discriminate inside-range stimulus amplitudes that they had previously judged from other nearby amplitudes. However, it is possible that the lower false alarm rates for outside-range distractors resulted from the larger target-distractor distance for those stimuli. Moreover, in the absence of explicit information to the contrary, it is possible that in this experiment the subjects believed that all distractors were outside the target amplitude range and thus only attempted to discriminate the outside-range distractors from all other stimuli. These defects were remedied in Experiment 3.

For Experiment 3, the proportion of subjects (proportion of trials for discrimination) who identified each stimulus as a previously judged target in the recognition and discrimination tests respectively was as follows (targets are in bold): 35 dB, .6 and .64; **40 dB**, .7 and .66; 45 dB, .6 and .69; **50 dB**, .9 and .69; 55 dB, .7 and .72; **60 dB**, .6 and .62; 65 dB, .2 and .58; **70 dB**, .6 and .59; 75 dB, .6 and .53; **80 dB**, .8 and .67; 85 dB, .9 and .63; **90 dB**, .4 and .63; and 95 dB, .1 and .07. In this experiment, subjects identified only the 95-dB distractor correctly as a distractor.

The d' analysis confirms this conclusion. Only the d' for the 90–95 pair was different from zero ($d' = 1.81, z = 7.54, p < .0001$). Inside-range d' values for 50-, 60-, 70-, and 80-dB targets and 45-, 55-, 65-, and 75-dB distractors ($d'_{lo} = 0.02$) and for 55-, 65-, 75-, and 85-dB distractors ($d'_{hi} = 0.05$) were also not significantly different from zero. Thus, there is no evidence in these data that subjects were able to recognize target stimuli as previously judged in the presence of distractors 5 dB from the various targets, unless the distractor was above the range of previously judged targets.

CONCLUSIONS

The present studies have revealed several important facts about peoples' memories of previous psychophysical scaling sessions. First, recalled psychophysical functions had the same slope as that of judged functions. However, subjects recalled veridically only their extreme responses and overestimated responses to intermediate amplitude stimuli, which caused recalled psychophysical functions to violate the power law. Second, people recognized the extreme stimulus amplitudes fairly well, usually rejecting both higher and lower amplitudes as not previously judged, and accepting all stimulus amplitudes between the extremes as ones that they had previously experienced. Moreover, they showed no evidence of being able to distinguish between previously experienced and not experienced stimulus amplitudes that lay between the extremes. Such a result would be surprising in an experiment with, for example, six target words and seven distractor words. Even if the distractors were repeated many times in a subsequent discrimination test and became as familiar as the targets, it seems apparent that the two sets of words would continue to be easily distinguished. It is possible that a more sensitive test, such as, for example, an implicit memory test employing a reaction time measure, might reveal some residual memory for the judged stimuli. In the present experiments, however, subjects managed to re-create from memory the slope of the S–R mapping established in a scaling session by remembering only two S–R pairings: those at the extreme high and low ends of the stimulus and response ensembles.

This tendency to remember only the extreme S–R pairings is consistent with theories that have proposed that people use the extreme stimuli as anchors and standards of judgment for responses to other stimuli (e.g., Braida & Durlach, 1988; Marley & Cook, 1986). It is reasonable to suppose that such use of extreme stimuli and their repeated pairing with roughly the same responses would enter them more firmly into long-term memory and make them easier to retrieve than the amplitudes and responses between those extremes, which would both be more variable (e.g., Braida & Durlach, 1988) and less often referred to. Since varying recall set had no effect on this tendency, subjects in the intentional memory conditions must have

strategically focused on their responses to extreme stimuli even when they were instructed to try to remember all of their responses to all of the stimuli.

These results also imply that the main assumption of the uncertainty hypothesis of mnemophysics seems to be correct: Memory is not precise for most of the stimuli that are used in a typical unidimensional experiment. Except for the extremes, the subjects were not able to discriminate stimuli that they had received during the scaling session from distractors displaced from the targets by 2.5 dB (Experiment 2) or by 5 dB (Experiment 3) within the same range. Remember that the differential threshold for pure tones in this intensity range is of the order of 0.5 dB, so that adjacent targets and distractors would have been easy to discriminate in a two-alternative forced-choice task. Braida and Durlach (e.g., 1988) attributed this disparity between ideal intensity resolution performance (the differential threshold) and performance in other situations (such as absolute identification) mostly to memory variance—in the present case, to an inability to remember the entire stimulus ensemble accurately (what they called “context coding mode”).

Second, there was no response range constriction in the present experiments. Rather, the responses to intermediate stimuli were overestimated. Moreover, Experiments 2 and 3 provided direct evidence that the memories of the stimuli themselves were vague. Both of these facts tend to support Algom et al.'s (1985) version of the uncertainty hypothesis. Third, the exponent of the power function for the remembered stimuli was identical to that for perceived stimuli. This result is not predicted by either of the two major mnemophysical hypotheses, although it has been found before for loudness of tones (Algom, 1992). Perhaps the relation between mnemophysical and psychophysical functions for intensive sensory continua, such as loudness, is different from that for extensive continua, such as line length. More research is needed in order to evaluate this conjecture.

Finally, these results have several implications for psychophysical scaling practice. In order to obtain “fresh” scaling data, uncontaminated by previous experience, it is necessary at the least to instruct subjects to disregard previous experiences. Various other disruptive techniques, such as changing the modulus (M. Teghtsoonian & R. Teghtsoonian, 1983) or using constrained scaling (West, Ward, & Khosla, in press) to establish the scale that is desired would also be useful to avoid biased scales. Perhaps most important is the implication that people always make scaling judgments within a context, that context includes memories of previous relevant experiences, and that the most relevant experience for any subject in a scaling experiment comprises previous scaling experiments. Thus, subjects are led to compromises between absolute and relative scales, and between memories and perceptions, in generating their responses to current stimuli (cf. Marks et al. 1986; Ward, 1987).

REFERENCES

- ALGOM, D. (1992). Memory psychophysics: An examination of its perceptual and cognitive prospects. In D. Algom (Ed.), *Psychophysical approaches to cognition* (pp. 441-513). Amsterdam: North-Holland.
- ALGOM, D., & LUBEL, S. (1994). Psychophysics in the field: Perception and memory for labor pain. *Perception & Psychophysics*, **55**, 133-141.
- ALGOM, D., & MARKS, L. E. (1989). Memory psychophysics for taste. *Bulletin of the Psychonomic Society*, **27**, 257-259.
- ALGOM, D., WOLF, Y., & BERGMAN, B. (1985). Integration of stimulus dimensions in perception and memory: Composition rules and psychophysical relations. *Journal of Experimental Psychology: General*, **114**, 451-471.
- BRAIDA, L. D., & DURLACH, N. I. (1988). Peripheral and central factors in intensity perception. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Auditory function* (pp. 559-583). New York: Wiley.
- GREEN, D. M., & SWETS, J. A. (1974). *Signal detection theory and psychophysics*. Huntington, NY: Krieger.
- HELSON, H. (1964). *Adaptation level theory: An experimental and systematic approach to behavior*. New York: Harper & Row.
- KERST, S. M., & HOWARD, J. H., JR. (1978). Memory psychophysics for visual area and length. *Memory & Cognition*, **6**, 327-335.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- MARKS, L. E. (1991). Reliability of magnitude matching. *Perception & Psychophysics*, **49**, 31-37.
- MARKS, L. E., SZCZESIUL, R., & OHLOTT, P. (1986). On the cross-modality perception of intensity. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 517-534.
- MARLEY, A. A. J., & COOK, V. T. (1986). A limited capacity rehearsal model for psychological judgments applied to magnitude estimation. *Journal of Mathematical Psychology*, **30**, 339-390.
- MOYER, R. S., BRADLEY, D. R., SORENSEN, M. H., WHITING, J. C., & MANSFIELD, D. P. (1978). Psychophysical functions for perceived and remembered size. *Science*, **200**, 330-332.
- NEISSER, U. (1981). John Dean's memory: A case study. *Cognition*, **9**, 1-22.
- SIEGEL, J. A., & SIEGEL, W. (1972). Absolute judgment and paired-associate learning: Kissing cousins or identical twins? *Psychological Review*, **79**, 300-316.
- TEGHTSOONIAN, M., & TEGHTSOONIAN, R. (1983). Consistency of individual exponents in cross-modal matching. *Perception & Psychophysics*, **33**, 203-214.
- VERRILLO, R. T. (1983). Stability of line-length estimates using the method of absolute magnitude estimation. *Perception & Psychophysics*, **33**, 261-265.
- 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. (1987). Remembrance of sounds past: Memory and psychophysical scaling. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 216-227.
- WEST, R. L., WARD, L. M., & KHOSLA, R. (in press). Constrained scaling: The effect of learned psychophysical scales on idiosyncratic response bias. *Perception & Psychophysics*.
- ZWISLOCKI, J. J., & GOODMAN, D. A. (1980). Absolute scaling of sensory magnitudes: A validation. *Perception & Psychophysics*, **28**, 28-38.

(Manuscript received June 18, 1998;
accepted for publication July 14, 1998.)