Learning in a unidimensional absolute identification task

JEFFREY N. ROUDER, RICHARD D. MOREY, NELSON COWAN, and MONIQUE PFALTZ
University of Missouri, Columbia, Missouri

We tested whether there is long-term learning in the absolute identification of line lengths. Line lengths are unidimensional stimuli, and there is a common belief that learning of these stimuli quickly reaches a low-level asymptote of about seven items and progresses no more. We show that this is not the case. Our participants served in a 1.5-h session each day for over a week. Although they did not achieve perfect performance, they continued to improve day by day throughout the week and eventually learned to distinguish between 12 and 20 line lengths. These results are in contrast to common characterizations of learning in absolute identification tasks with unidimensional stimuli. We suggest that this learning reflects improvement in short-term processing.

The goal of this paper is to document that people learn in an absolute identification task with unidimensional stimuli. We provide clues as to how this learning is possible despite the conventional wisdom that learning does not happen in these tasks with these stimuli. Under ordinary circumstances, learning is not surprising. People learn to identify new people, places, objects, and words. Our ability to learn can be characterized as *unbounded*, in that if we devote a sufficient amount of time and effort, we can learn an arbitrary number of new items to a very high degree of precision. For example, high school students may learn hundreds of new SAT vocabulary words, medical students may learn hundreds or thousands of new anatomical terms, and college professors eventually learn students' names. Although the rates of learning vary, there appears to be no known magic number that limits the amount we can learn. In the extreme, the conjecture of unbounded learning may be false, since we are material and finite. However, this is irrelevant; to first approximation, the notion of unbounded learning serves as a useful characterization for many domains.

There are domains, however, in which the unbounded learning conjecture fails dramatically. One much-studied example is that of unidimensional stimuli—stimuli that vary on a single physical dimension. Examples of unidimensional stimuli include lines of varying length and tones of varying intensity. Throughout the 1950s, researchers documented that people were exceedingly poor at learning unidimensional stimuli, often learning no more than about seven items (Hake & Garner, 1951; Miller, 1956; Pollack, 1952; Shiffrin & Nosofsky, 1994).

The prevailing characterization is that with increasing practice, performance quickly reaches a low-level asymptote and improves no more (Lacouture, Li, & Marley, 1998; Shiffrin & Nosofsky, 1994). For example, Shiffrin and Nosofsky wrote:

> Surely, one imagines, if one spends enough time rehearsing and practicing, perfect performance should eventually be achieved. Such an effect has never been demonstrated in the experimental literature, however. As an anecdotal example, Robert M. Nosofsky started his research career around 1980 in the acoustical laboratory of David Green and R. Duncan Luce, two researchers who happened at the time to be studying absolute identification of loudness (among numerous other discrimination and magnitude estimation processes). As a cocky young graduate student, Nosofsky "knew" that with a bit of practice, he could surely learn to perfectly identify a set of twelve loudnesses. After locking himself in one of Green's sound-proof booths for several weeks, and hearing tone after tone after tone, his absolute identification performance remained unchanged. He did succeed, however, in increasing substantially his need for psychotherapy. (p. 357)

This prevailing characterization of severe learning deficits in unidimensional stimuli concords with a modal view of memory and identification. Miller (1956) wedded short-term memory and absolute identification in his classic paper about the magic number seven. Seven described the capacity of chunks in short-term memory tasks as well as the number of unidimensional stimuli that can be distinguished in absolute identification tasks. On the basis of this numeric commonality, Miller raised the possibility that there is a common capacity-limited mechanism underlying both domains. This conjecture of a common mechanism can be combined with Atkinson and Shiffrin's (1968) distinction between short- and long-term memory. In this modal view, unidimensional stimuli are not stored in long-term memory. Instead, they can only be held in short-term memory (e.g., Marley &

Cook, 1984). Poor unidimensional absolute identification performance directly reflects the capacity limits of short-term memory. With these stimuli, there is no long-term improvement with practice because there is no opportunity to store stimulus-specific information (i.e., the perceived value of a stimulus on a relevant dimension) in long-term memory. This inability to store stimuli in long-term memory is unique to unidimensional stimuli. More complex relationships can be stored and used subsequently in object recognition. An example of an object recognition theory that postulates representation of only multidimensional primitives is Biederman's (1987) recognition-by-components theory.

In accordance with this modal view, we expected performance to reach asymptote within a hundred trials or so. We found, however, that this was not the case. Our well-motivated participants were still improving after thousands of trials of practice. We explain our results with a revision of the modal model in which practice improves short-term processing. Our results further suggest that this improvement reflects improved encoding of the length of the stimulus rather than rote memorization of stimulus–response associations. Our results with absolute identification are analogous to those of Ericsson, Chase, and Faloon (1980) with immediate serial recall, which indicate that extended practice raises digit span from 7 random digits to 79. In both cases, learning reflects improved short-term processing rather than rote memorization of the entire stimulus set. It is in this type of improvement, rather than in Miller's (1956) magic number 7, that we found commonality between the task domains.

METHOD

Participants

Participants were two of the authors (R.M. and M.P.) and an unaffiliated volunteer who was monetarily compensated for his time (D.M.). All 3 participants were highly motivated and took the challenge of learning the sets of line lengths with the utmost seriousness and determination.

Stimuli

Three sets of line lengths were constructed (Table 1). The small, medium, and large sets consisted of 13, 20, and 30 line lengths, re-

Table 1
Line Lengths in Pixels

Small Set									
15	23	32	43	57	73	93	116	143	174
210	251	298							
Medium Set									
9	12	16	20	26	33	41	50	60	72
85	100	117	136	157	180	206	234	264	298
Large Set									
9	12	14	17	20	23	27	31	36	41
47	53	60	67	76	84	94	104	115	127
140	153	168	183	199	217	235	255	276	298

spectively. In all of the sets, increments between line lengths increased as a power function (Stevens, 1975) with an exponent of 3.5. Figure 1 depicts line lengths in the large set. The 17-in. display was programmed to 640 × 480 pixel mode. Each pixel measured .5 mm, and the participants sat approximately 50 cm from the display. To ensure that the edge of the monitor could not be used as a relative size cue, we varied the location of the line center, which was chosen randomly from a uniform distribution over the 22 × 22 pixel center-most box. With this random variability, the distance between the line endpoints was the only reliable indicator of length.

Design and Procedure

The experiments were absolute identification tasks in which the smallest line of the set was paired to the numerical label "1," the next to "2," and so on. A trial began with a blank screen. After a 500-msec foreperiod, a line length was displayed and remained present until the end of the trial. The participant depressed the key with the chosen numerical label. For correct responses, the participant received a pleasant-sounding sequence of a 500-hz tone followed by a 1000-Hz tone. For incorrect responses, the participant saw the message "Try again," and responded again. If this second response was correct, the participant received the pleasant-sounding sequence. Otherwise, the correct response number was displayed under the line for 500 msec. Second responses were allowed to increase participants' perceived performance levels, and hence their overall morale. Unless indicated, analyses are from the first response only. Each block consisted of 65 trials (Experiment 1) or 72 trials (Experiments 2 and 3). Preceding each block was a display similar to that of Figure 1, showing all of the stimulus–response pairings. Following a block, participants took a minute-long break. Sessions, which were performed on consecutive weekdays, comprised 10 blocks of trials. Sessions lasted about 90 min.

Experiment 1. Participants R.M. and M.P. identified the small set of 13 line lengths for 5 and 10 consecutive sessions, respectively. Participant R.M. reached ceiling after 5 sessions, prompting early termination of the experiment. He returned 9 days later to complete an additional 6th session. The use of a 9-day retention interval allows us to assess the medium-term retention of learning.

Experiment 2. After taking part in Experiment 1, Participants R.M. and M.P. identified the medium set of 20 line lengths for 4 and 10 sessions, respectively. Once again, the experiment was terminated because of R.M.'s early ceiling-level performance.

Experiment 3. After taking part in Experiment 2, Participants R.M. and M.P. identified the large set of 30 line lengths for six and seven sessions, respectively. Participant D.M. identified the large set for seven sessions.
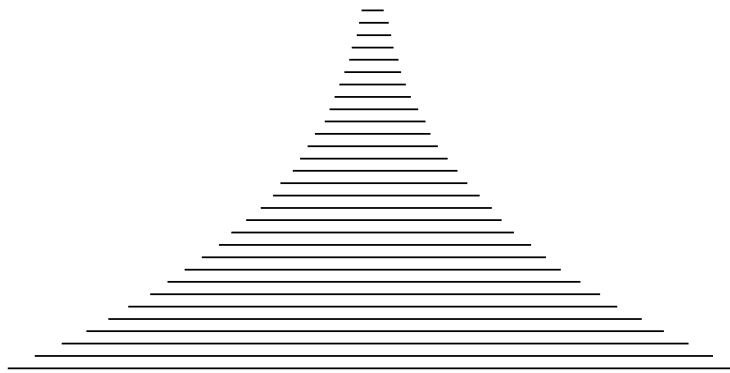
RESULTS

Results are provided in four separate analyses: (1) overall learning, (2) serial position curves, (3) savings across rest intervals, and (4) trial-by-trial sequential effects. These analyses together document learning and suggest a locus for the effect.

Overall Learning

The main result obtained for all participants in all experiments is that of improvement from session to session. The left column of Figure 2 shows accuracy as a function of session for the first response. Improvements in overall session accuracy varied depending on the initial accuracy. For Experiment 3, in which all 3 participants had a moderate initial accuracy, the average per-

Figure 1. Line lengths in the large set of 30 stimuli. Lengths were computed by a power function with an exponent of 3.5.

formance improved gradually from .42 to .70, a gain of .28. The open point in the top panel is from the delayed session for Participant R.M. He maintained his high level of identification performance even after a 9-day delay.

An alternative measure to percent correct is the number of transmitted bits per trial (Shannon, 1948). The middle column of Figure 2 shows bits transmitted as computed from the confusion matrices (Garner & Hake, 1951). The middle column shows transmitted bits on a log scale. Because transmitted bits constitute a log measure, it is convenient to use 2^{bits} as a performance index. This index, termed *equivalent correct*, represents the number of stimuli that could be perfectly identified. In
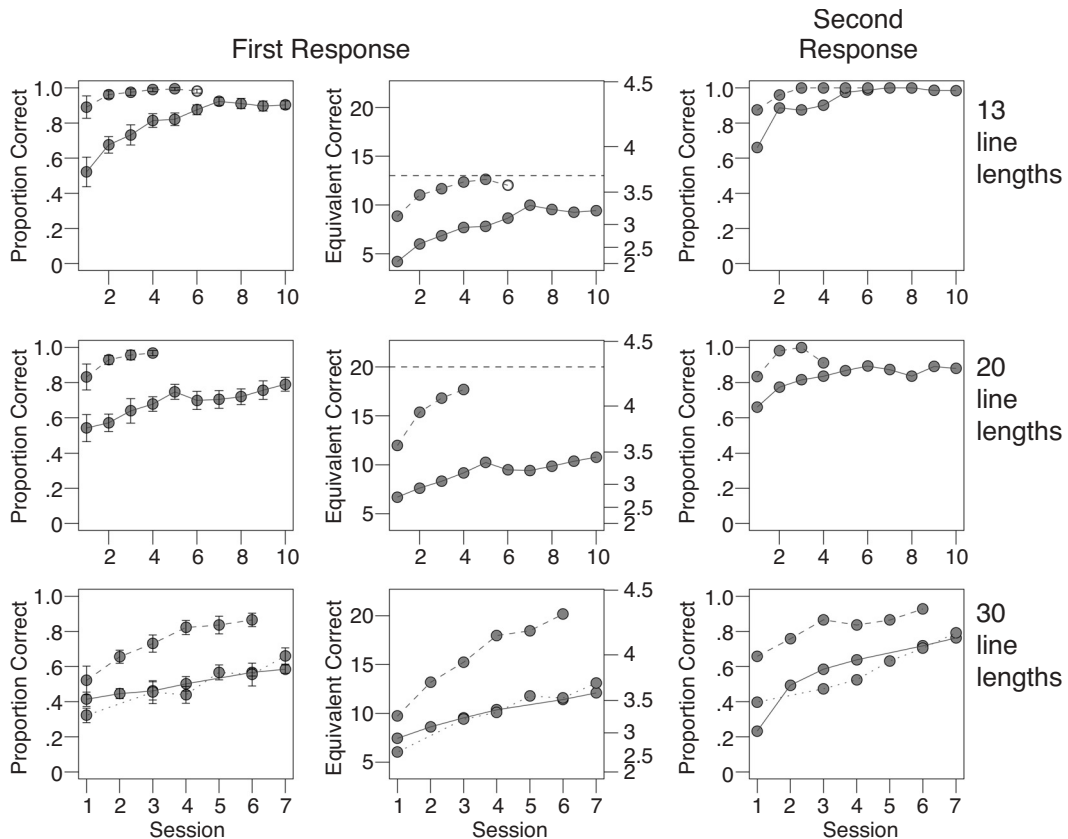
Figure 2. Accuracy and transmitted bits as a function of session in the three experiments. The left and center columns represent analyses of the first response; the right column shows analysis of the second response when the first one was incorrect. Solid, dashed, and dotted lines correspond to data from participants M.P., R.M., and D.M., respectively. The open circle in the top plots is from a session that occurred 9 days after the previous one. Error bars denote 95% confidence intervals.

Miller (1956) the magic number 7, as applied to absolute identification, refers to the equivalent correct measure. The left-hand axis of the middle column shows performance as measured with the equivalent correct index. Finally, the right-hand column shows performance on the second response after an incorrect first response. Learning is evident here too.

Serial Position Curves

Serial position curves are plotted in Figure 3. As is typical, performance was better in the extremes than in the middle. There was more learning in the middle of the range than at the extremes, although the effect for the smallest line lengths is difficult to interpret because of near-ceiling-level performance in all participants.

Although performance gains are in the middle ranges, it is difficult to rule out a ceiling effect for the more extreme stimuli.

Loss and Recovery

Given that our participants learned on a long-term basis, we assessed whether there was loss when participants were not performing the task. There were two such intervals, the 1-min breaks between blocks and the much larger intervals between sessions. Figure 4 shows performance in Experiment 3 on a trial-by-trial basis. The top panel shows performance before and after a 1-min break; the bottom panel shows performance before and after an intersession interval. The dashed line connects the data points, and the solid line smooths them (LOWESS algorithm, Cleveland, 1981).

In the top panel, performance on the first trial following a break is indicated with an arrow. Accuracy is lower (.50 vs. .58 baseline) but recovers within the next few trials. The first trial is the unique minimum, and the prob-

ability of this happening under the considered null hypothesis is 1 in 72, less than .05. There is a similar pattern following an intersession interval (first trial performance of .33 vs. a .55 baseline), but it appears to be within sampling variability. Of course, the bottom graph suffers from low power (there were five intersession intervals and 54 breaks per participant). Overall, it appears that there may be losses associated with breaks, but recovery to baseline is much faster than initial learning, indicating a savings.

Sequential Effects

A possible locus for the learning phenomenon is suggested by the following analysis of trial-by-trial sequential effects. The existence of these effects is well documented in absolute identification (Lockhead & King, 1983; Luce, Nosofsky, Green, & Smith, 1982; Mori, 1998; Treisman, 1985; Ward & Lockhead, 1970). We define the error of a response as the difference between the response and the ordinal value of the stimulus. It is well known that a previous stimulus induces an assimilative bias—the response is biased in the direction of the previous stimulus. Figure 5 shows an analysis of sequential effects. The top panel shows mean error as a function of the ordinal distance between the current stimulus and the previous one. The overall finding of positive error for negative differences and negative error for positive differences indicates assimilation. The magnitude of these bias effects was unaffected by practice (solid lines denote performance averaged over the second and third sessions; dashed lines denote performance averaged over the fifth and sixth sessions).

Mean error is only one type of sequential-effect measure; we also document an effect in squared error. If the previous stimulus was close to the current one (and a little smaller than the current one), then error was reduced in magnitude and performance was improved (center and bottom panels of Figure 5). This trend is especially evident for the first few days and is similar to but stronger in magnitude than the one documented in by Luce et al., 1982. Increased practice affects performance overall, but it especially aids performance when the previous stimulus is far from the current one. One explanation for these findings is that the current trial is calibrated from the previous one. Calibrating from nearby stimuli is initially easier than from faraway stimuli, but with practice participants learn to better calibrate from faraway ones too. The effect of calibrating may be to add some assimilative bias but, more importantly, to reduce the variance of responses, and as a consequence, improve performance.

DISCUSSION

The preceding analyses show that although none of our participants achieved perfect levels of performance, all continued to learn for thousands of trials in the absolute identification of unidimensional stimuli. In the introduction, we noted that this is not the result expected in the field, and it directly conflicts with characteriza-
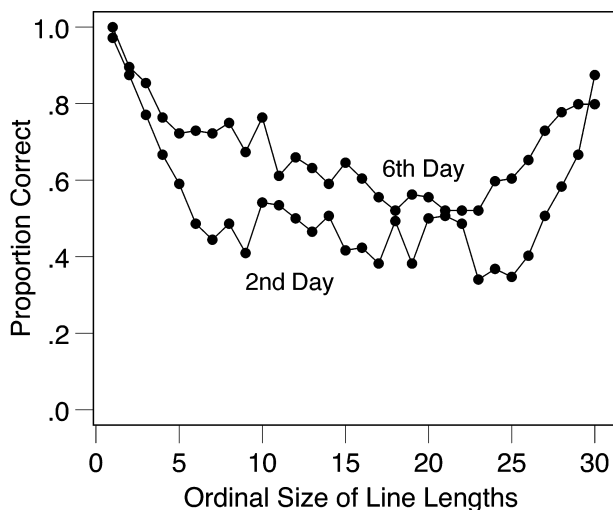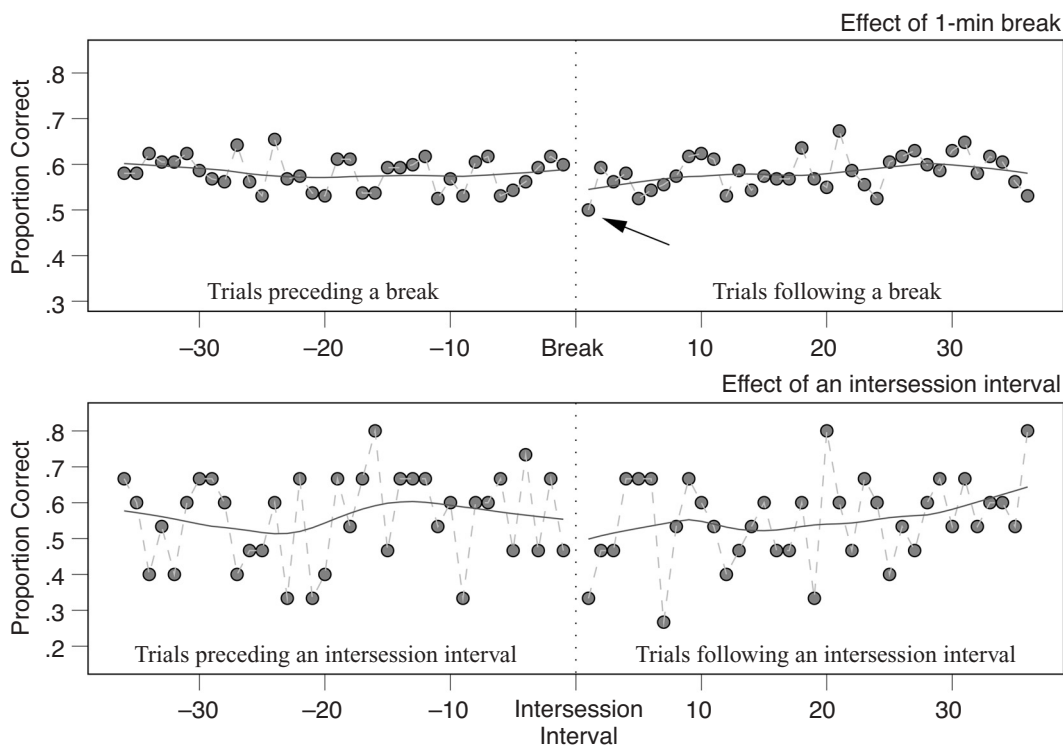
Figure 3. Averaged proportion correct as a function of stimulus length (ordinal values) for Experiment 3. The lower line shows performance on the 2nd day; the upper line shows the same on the 6th day.

Figure 4. Averaged proportion correct as a function of trial for Experiment 3. Proportions were averaged across blocks, participants, and days (except for the 1st day).

tions by Lacouture et al. (1998) and Shiffrin and Nosofsky (1994). Other authors have at least implicitly assumed there was no learning when aggregating their data across hundreds, even thousands, of trials (e.g., Gravetter & Lockhead, 1973; Hake & Garner, 1951; Luce et al., 1982; Ward & Lockhead, 1970). Perhaps they did examine learning and found none, or perhaps they aggregated without examination. In either case, none of these authors comments on the possibility of learning. Perhaps this oversight in the field results from an implicit comparison of performance to that with more complex, multidimensional stimuli. The learning documented here is painfully slow, especially in comparison with the rapid learning of words and objects. In comparison with the learning of these more complex stimuli, a lack of learning is a coarse approximation to our data. We are aware of only one other study that documents learning with unidimensional stimuli. Weber, Green, and Luce (1977) collected 12,000 observations per participant. Performance on the last 2,000 of these was about seven percentage points higher than that on the first 2,000.

Absolute identification of pitch is a related domain in which learning has been documented (see Takeuchi & Hulse, 1993). Pitch, however, is not analogous to line lengths because the psychological representation of pure tones is thought to be multidimensional, with octaves being closer than, say, sixths (Shepard, 1982). Perception of visual length and auditory intensity, on the other hand, is assumed to correspond to unidimensional psy-

chological correlates (Durlach & Braida, 1969; Gravetter & Lockhead, 1973; Karpiuk, Lacouture, & Marley, 1997; Luce, Green, & Weber, 1976; Marley & Cook, 1984).

Although it is important to establish an accurate characterization of the results, it is equally important that they be used to address theory. We started this paper with the presentation of a modal model. Identification of unidimensional stimuli suffered from the same capacity limit as short-term memory because stimulus–response associations between unidimensional stimuli and assigned responses could not be formed or maintained in long-term memory. Almost all models of unidimensional identification subscribe, at least implicitly, to this modal view (e.g., Durlach & Braida, 1969; Gravetter & Lockhead, 1973; Lacouture et al., 1998; Luce et al., 1976; Marley & Cook, 1984; Treisman, 1985).

One alternative is to assume participants do have recourse to encoding specific stimulus–response paired associations (e.g., Haubensak, 1992; Siegel & Siegel, 1972). As practice occurs, participants would transition from a magnitude estimation strategy to one that is mediated by long-term stimulus–response associations. However, this view does not explain some of the interactions in the results: the observed sequential effects, which do not change with practice in terms of mean error, or the much greater improvement in performance on a particular line length when it is preceded by a dissimilar line length as opposed to a similar one. Further-
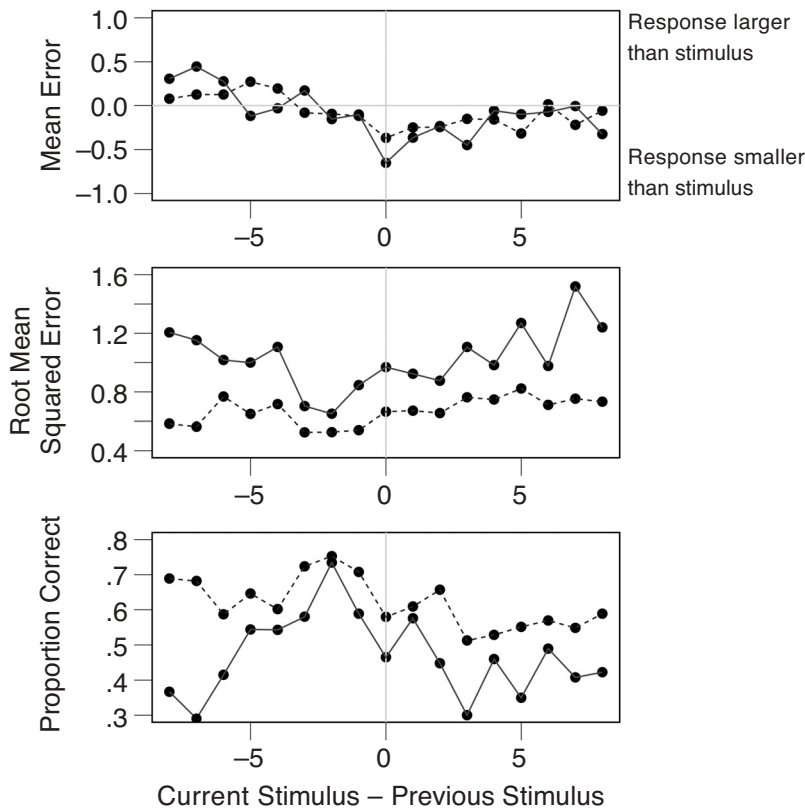
Figure 5. Top, center, and bottom panels show averaged mean error, root mean squared error, and proportion correct, respectively, as a function of the distance between current and previous stimuli in Experiment 3. The solid and dashed lines indicate performance for earlier sessions (2 and 3) and later sessions (4 and 5), respectively. Note that there was no practice effect in mean error (i.e., no change with practice in assimilation bias), yet there was an interaction between practice and distance in root mean square error (i.e., decrease with practice in the variability of response on trials with large distances).

more, a long-term storage theory does not provide an explanation of the vastly different learning rates between simple and complex stimuli.

The learning result can be explained better with a relatively minor revision of the modal model. According to most theorists, performance on a trial involves what we term "calibrated magnitude estimation." The basic idea is that participants estimate the line length and then map these estimates onto responses through a set of decision bounds. In some theories, bound placement is thought to vary on a trial-by-trial basis (Gravetter & Lockhead, 1973; Treisman, 1985). To explain the observed results, we hypothesize that without practice, the bounds are variable when the previous stimulus is far from the current one. However, with practice, participants are better at using the previous stimulus for calibrating responses for both nearby and faraway stimuli. This explanation, while speculative, is concordant with current models of unidimensional absolute identification (e.g., Marley & Cook, 1984; Purks, Callahan, Braida, & Durlach, 1980; Treisman, 1985).

This view of practice affecting short-term abilities is consistent with the seminal results of Ericsson et al. (1980), who found a remarkable improvement in short-term capacity, from 7 to 79 digits, with 230 hours of practice over 20 months. None of this increase can have been due to long-term retrieval of specific digit sequences because these were randomly chosen and always changing. Instead, Ericsson et al. argued that the tremendous gain was due to the development of on-line encoding skills. Their participant related sequences to plausible times it takes to run standard distances (e.g., 3,492 can be coded as 3 minutes, 49.2 seconds, a near-record time for the mile) or to ages (e.g., 893 can be coded as 89 years and 3 months, a very old man). The Ericsson et al. result shows the plausibility of practice improving short-term processes rather than implicating the existence of a long-term storage of specific stimulus–response pairings. Their dramatic demonstration raises the possibility that our participants could learn much more if they participated for 20 months. However, in the passage cited in the introduction, Shiffrin and Nosofsky

(1994) wrote that listening to the same stimuli for several weeks substantially increased Nosofsky's need for psychotherapy. We fear that 20 months of line-length identifications would place participants *beyond the reach* of psychotherapy. More likely, though, participants would stop trying to improve and might therefore stop improving, as it has been argued that expertise in a new skill depends not merely on extended practice, but on deliberate practice, in which participants make unflagging attempts to improve performance strategies (Ericsson & Krampe, 1993).

In sum, our data are novel in that we document slow but steady learning in an absolute identification task with unidimensional stimuli. Practice results in a disproportionate improvement in performance when the previous stimulus is far from the current one. This context dependence of learning suggests to us that improvement occurred in short-term encoding abilities rather than long-term stimulus–response mappings, although the issue cannot be fully resolved with our data. We hope that comparisons between rates and loci of learning effects across different stimulus classes will provide continuing insight into the nature of human capacities and their limitations.

REFERENCES

ATKINSON, R. C., & SHIFFRIN, R. N. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.

BIEDERMAN, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, **94**, 115-147.

CLEVELAND, W. S. (1981). LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *American Statistician*, **35**, 54.

DURLACH, N. I., & BRAIDA, L. D. (1969). Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, **46**, 372-383.

ERICSSON, K. A. CHASE, W. G., & FALOON, S. (1980). Acquisition of a memory skill. *Science*, **208**, 1181-1182.

ERICSSON, K. A., & KRAMPE, R. T. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, **100**, 363-406.

GARNER, W. R., & HAKE, H. W. (1951). The amount of information in absolute judgments. *Psychological Review*, **58**, 446-459.

GRAVETTER, F., & LOCKHEAD, G. R. (1973). Criterial range as a frame of reference for stimulus judgement. *Psychological Review*, **80**, 203-216.

HAKE, H. W., & GARNER, W. R. (1951). The effect of presenting various numbers of discrete steps on scale reading accuracy. *Journal of Experimental Psychology*, **42**, 358-366.

HAUBENSAK, G. (1992). The consistency model: A process model for absolute judgments. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 303-309.

KARPIUK, P., JR., LACOUTURE, Y., & MARLEY, A. A. J. (1997). A limited capacity, wave equality, random walk model of absolute identification. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 329-346). Mahwah, NJ: Erlbaum.

LACOUTURE, Y., LI, S., & MARLEY, A. A. J. (1998). The roles of stimulus and response set size in the identification and categorization of unidimensional stimuli. *Australian Journal of Psychology*, **50**, 165-174.

LOCKHEAD, G. R., & KING, M. C. (1983). A memory model of sequential effects in scaling tasks. *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 461-473.

LUCE, R. D., GREEN, D. M., & WEBER, D. L. (1976). Attention bands in absolute identification. *Perception & Psychophysics*, **20**, 49-54.

LUCE, R. D., NOSOFSKY, R. M., GREEN, D. M., & SMITH, A. F. (1982). The bow and sequential effects in absolute identification. *Perception & Psychophysics*, **32**, 397-408.

MARLEY, A. A. J., & COOK, V. T. (1984). A fixed rehearsal capacity interpretation of limits on absolute identification performance. *British Journal of Mathematical & Statistical Psychology*, **37**, 136-151.

MILLER, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, **63**, 81-97.

MORI, S. (1998). Effects of stimulus information and number of stimuli on sequential dependencies in absolute identification. *Canadian Journal of Experimental Psychology*, **52**, 72-83.

POLLACK, I. (1952). The information of elementary auditory displays. *Journal of the Acoustical Society of America*, **24**, 745-749.

PURKS, S. R., CALLAHAN, D. J., BRAIDA, L. D., & DURLACH, N. I. (1980). Intensity perception. *Journal of the Acoustical Society of America*, **67**, 634-637.

SHANNON, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379-423.

SHEPARD, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, **89**, 305-333.

SHIFFRIN, R. M., & NOSOFSKY, R. M. (1994). Seven plus or minus two: A commentary on capacity limitations. *Psychological Review*, **101**, 357-361.

SIEGEL, J. A., & SIEGEL, W. (1972). Absolute judgment and paired-associate learning: Kissing cousins or identical twins? *Psychological Review*, **79**, 300-316.

STEVENS, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.

TAKEUCHI, A. H., & HULSE, S. H. (1993). Absolute pitch. *Psychological Bulletin*, **113**, 345-361.

TREISMAN, M. (1985). The magical number seven and some other features of category scaling: Properties of a model for absolute judgment. *Journal of Mathematical Psychology*, **29**, 175-230.

WARD, L. M., & LOCKHEAD, G. R. (1970). Sequential effects and memory in category judgments. *Journal of Experimental Psychology*, **84**, 27-34.

WEBER, D. L., GREEN, D. M., & LUCE, R. D. (1977). Effects of practice and distribution of auditory signals on absolute identification. *Perception & Psychophysics*, **22**, 223-231.

NOTES

1. The exponent of 3.5 was chosen as follows. We picked the smallest and largest stimuli a priori. We then experimented with various exponents for the small set of 13 stimuli and found that with a value of 3.5, R.M. and M.P. could discriminate the smallest from the second-smallest stimuli in a discrimination task with 100% accuracy. They could do the same with the largest and second-largest stimuli.

2. The sequential effects analysis is based on Stimuli 9–22 and distances between the current and previous stimulus from −8 to 8. This restriction ensures performance at each level of distance was based on the same stimuli.