

Increasing Capacity: Practice Effects in Absolute Identification

Pennie Dodds
University of Newcastle

Christopher Donkin
University of Indiana

Scott D. Brown and Andrew Heathcote
University of Newcastle

In most of the long history of the study of absolute identification—since Miller’s (1956) seminal article—a severe limit on performance has been observed, and this limit has resisted improvement even by extensive practice. In a startling result, Rouder, Morey, Cowan, and Pfaltz (2004) found substantially improved performance with practice in the absolute identification of line lengths, albeit for only 3 participants and in a somewhat atypical paradigm. We investigated the limits of this effect and found that it also occurs in more typical paradigms, is not limited to a few virtuoso participants or due to relative judgment strategies, and generalizes to some (e.g., line inclination and tone frequency) but not other (e.g., tone loudness) dimensions. We also observed, apart from differences between dimensions, 2 unusual aspects of improvement with practice: (a) a positive correlation between initial performance and the effect of practice and (b) a large reduction in a characteristic trial-to-trial decision bias with practice.

Keywords: absolute identification, practice, capacity, unidimensional

Human memory for complex items such as names, letters, and faces is seemingly infinite. People are able to memorize a great number of these items across their life span, or even in a 1-hr experimental task, with relative ease. For decades, however, a single, simple task has provided an exception to this rule: absolute identification. Absolute identification (sometimes also referred to as dead reckoning) is the task of identifying which stimulus has been shown out of a set of stimuli that vary on only one physical dimension. For example, a participant might be given a set of n lines varying in length, or tones varying in intensity, labeled 1 through n . On each trial of an absolute identification task, the participant is then presented with one of these stimuli and asked to recall its label. Empirical research into absolute identification has a long history, with Miller’s (1956) summary of early work identifying a surprisingly small capacity limitation—people are generally unable to accurately identify more than 7 ± 2 stimuli in an absolute identification task. Miller noticed a similar limitation in short-term memory performance, and the two limitations have often been treated as manifestations of a single phenomenon; that is, absolute identification performance is limited precisely because

it relies on short-term memory capacity, and so the study of absolute identification is interesting (in part) because of what it reveals about short-term memory.

For decades, the received view has been that this capacity limitation is unaffected by manipulations that are otherwise very powerful. For example, Miller (1956) showed that the capacity limit was about the same for the absolute identification of many different kinds of stimuli, including line length, taste, brightness, hue, and loudness. There are many other stimulus manipulations that one might assume would improve performance, but these have all been demonstrated to have little or no effect on the capacity limitation (e.g., increasing the number, separating the stimuli; Garner, 1953; Pollack, 1952). Possibly the most intriguing finding is that the capacity limit is highly resistant to practice. For example, Garner’s (1953) participants engaged in up to 12,000 judgments in a single condition, yet even at the end of the experiment they were still limited to identifying the equivalent of three or four stimuli correctly. Weber, Green, and Luce (1977) had participants complete 12,000 trials identifying six white noise signals of varying loudness and found an improvement in response accuracy of just 6%. Final performance for these participants was well below ceiling, despite the large amount of practice, monetary incentives, and the apparently easy task of identifying just six separate levels of loudness. Hartman’s (1954) participants also practiced over an 8-week period, and although they demonstrated substantial improvement, their best performance level was still well within Miller’s limit—equivalent to the perfect identification of only five stimuli. Such results have established this truism about absolute identification: There is a severe limitation in human ability to identify unidimensional stimuli, and this limit is largely unaffected by practice.

In a departure from previous findings, Rouder, Morey, Cowan, and Pfaltz (2004) demonstrated that substantial learning *is* possible

This article was published Online First February 7, 2011.

Pennie Dodds, Scott D. Brown, and Andrew Heathcote, School of Psychology, University of Newcastle, Callaghan, New South Wales, Australia; Christopher Donkin, Department of Psychology and Brain Sciences, University of Indiana.

Parts of the research reported herein contributed to a doctoral thesis for Pennie Dodds. This research was supported, in part, by Australian Research Council Discovery Project DP0881244 to Scott D. Brown and Andrew Heathcote.

Correspondence concerning this article should be addressed to Pennie Dodds, School of Psychology, University of Newcastle, Callaghan, New South Wales 2308, Australia. E-mail: pennie.dodds@newcastle.edu.au

in an absolute identification task. In particular, three participants showed large improvements in the identification of line length with practice. One participant, after 11,100 trials of practice, was able to correctly identify almost 20 different line lengths. The other two participants, with 18,740 and 5,040 trials of practice, were able to correctly identify about 13 lines. It is not clear what caused the difference between Rouder et al.'s result and earlier studies. For example, learning may have been improved because Rouder et al.'s participants were given chances to correct incorrect responses. Perhaps also the large improvement with practice is unique to the absolute identification of line lengths and would not have been observed with, for example, the identification of tones of varying loudness (consistent with Garner's 1953, results). This explanation seems especially attractive because, although line lengths have been used occasionally in the field (see e.g., Kent & Lamberts, 2005; Lacouture, 1997; Rouder, 2001; Thorndike, 1932), previous demonstrations of the null effect of practice have mainly used tones varying in intensity. Another important difference between Rouder et al.'s methods and earlier work was the use of considerably larger stimulus sets (up to 30 different lines, rather than the more typical 8–12 stimuli).

These findings are particularly interesting because they might shed light on the deeper issue: Although people seem to have practically unlimited memory for items such as faces and names, unidimensional stimuli have been highlighted as the exception to this rule. Through a series of experiments, we investigate whether unidimensional stimuli truly represent an exception to this short-term memory limitation and what characteristics of such stimuli affect overall learning. As well as identifying which kinds of stimulus sets support learning and which do not, we also investigate the mechanisms underlying improvement with practice. For example, participants may learn to increase the capacity of their short-term memory and so are better able to pair stimuli with their to-be-recalled labels. Alternatively, they may learn to avoid some of the well-documented decision biases that pervade absolute identification (the "sequential effects"; see e.g., Stewart, Brown, & Chater, 2005, p. 883). To foreshadow our conclusions, although our data strongly suggest improvements of the latter variety, model-based analyses implicate both kinds of learning.

Experiment 1

We began our investigations by examining whether any of the atypical design features used in Rouder et al.'s (2004) study contributed to the large learning effect. The most novel aspect of Rouder et al.'s design was their response technique, in which participants were given two opportunities to respond instead of the

standard single response. If participants made an incorrect response, they were allowed a second attempt. If they were incorrect on their second attempt, the correct answer was displayed. In Experiment 1a, we aimed to replicate Rouder et al.'s findings of significant learning with their response method—to ensure that Rouder et al. did not simply have an exceptional sample of participants. In Experiment 1b we investigated whether learning persists with a standard response method in a paradigm that is otherwise identical.

Method

Participants. Six participants took part in Experiment 1a, and a different six took part in Experiment 1b. Each was reimbursed \$15 per session, and unless otherwise stated, this was the case in all following experiments, with six new participants recruited for each except Experiment 2, in which five participants were recruited for each condition.

Stimuli. The stimuli were 30 lines of varying length, increasing in size according to a power function with an exponent of 3.5 (see Rouder et al., 2004, and see Table 1). Stimuli were presented in black on a white background, using a 21-in. cathode ray tube monitor set at a resolution of $1,152 \times 864$ pixels. Each pixel measured 0.39 mm wide \times 0.35 mm high. Images were positioned in the center of the screen, with a 22×22 pixel variation in position from trial to trial to discourage participants from using the edge of the monitor as a size cue.

Procedure. In a brief study phase at the beginning of each session, participants were given each stimulus one at a time, labeled with a corresponding number, from 1 to 30. In order to proceed through the study phase, the participant had to select the number on-screen that corresponded with the numerical label. For example, an on-screen prompt would display *This is line number 1. Press 1 to continue.* During each trial in the main phase of the experiment, one stimulus was randomly selected and presented, and the participant was asked to respond with the numerical label that was attached to the stimulus in the study phase. Instructions given to the participants emphasized response accuracy over response time. This decision was made in light of our primary interest in how accurately participants could perform the task. Responses were made using the mouse to click buttons arranged on-screen in increasing numerical order. Three columns of 10 buttons were arranged on the left-hand side of the screen, and these remained on-screen throughout the experiment.

The only difference between Experiments 1a and 1b was the number of response opportunities per trial. In Experiment 1a, we replicated Rouder et al.'s (2004) two-response method. If partici-

Table 1
Line Lengths in Pixels for Experiments 1a, 1b, and 5a

Experiments 1a and 1b									
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
Experiment 5a									
15	18	22	27	33	41	50	61	74	90
110	134	164	200	244	298				

pants were incorrect on the first response, they were given a second response opportunity. If they were incorrect again, the correct answer was displayed for 500 ms. Whenever a correct response was recorded, the word *Correct* was displayed and the trial ended. In Experiment 1b, we used the traditional one-response absolute identification feedback system, in which participants were given only one opportunity to respond. If they were incorrect, the correct answer was displayed for 500 ms. If they were correct, the word *Correct* was displayed. The stimulus always remained on-screen until the final feedback was provided.

Participants took part in 10 sessions, each of which lasted approximately 1 hr. Sessions were conducted on (mostly) consecutive days. The first three sessions consisted of six blocks of 90 trials, whereas the remaining seven sessions consisted of seven blocks of 90 trials. This resulted in 201 presentations per stimulus per participant. A minimum 1-min break between blocks was enforced.

Results

Analyses were conducted on the first response only, to allow more valid comparison of Experiments 1a and 1b (cf. Rouder et al., 2004). Rather than focus on only response accuracy, we also calculated the amount of information transmitted from the stimulus to the response. Due to a historical focus on information-theoretic accounts of absolute identification performance (see e.g., Garner & Hake, 1951), information transfer has become a standard descriptor for performance, and it is also particularly useful when comparing different stimulus set sizes (see Shannon, 1948; also Garner, 1953; Pollack, 1952). Information transfer attempts to measure how much uncertainty in the identity of the stimulus is removed by considering the observer's response. The amount of information transmitted from the stimulus to the response is measured in "bits," and 2^{bits} can be interpreted as the equivalent number of stimuli that could be perfectly identified (e.g., 3 bits of transmitted information corresponds to perfectly accurate identification of $2^3 = 8$ stimuli).

We calculated the amount of transmitted information separately for each participant and each practice session. We quantified the amount of improvement induced by practice using the minimum and maximum of these values. Note that these extremes did not always occur in the first or last sessions, but analyses based on the first and last sessions yielded similar results. We employed the minimum and maximum values because of a trend for participants to lose some motivation in the final session of the experiment—across all experiments and all participants, the proportion of increases in performance from one session to the next was 72%, but this was significantly lower for the final session, at 41% ($\chi^2 = 4.2$, $p < .05$).

Improvement with practice was apparent in both Experiments 1a and 1b, as illustrated in Figure 1. In Experiment 1a, in which participants were given two response opportunities, the percentage of correct responses (i.e., accuracy) improved from 23% to 49%, compared with a chance performance level of just 3.3%. In terms of information transmission, this corresponds to an average improvement of 0.83 bits, from 2.40 to 3.23. This meant that average maximum performance (across participants) was equivalent to the perfect identification of approximately 9.4 stimuli. A one-way (session) repeated measures analysis of variance with a

Greenhouse–Geisser sphericity correction confirmed that these effects were highly reliable for both accuracy, $F(1.36, 6.63) = 14.84$, $p = .005$, and information transfer, $F(1.37, 6.72) = 26.17$, $p = .001$.

In Experiment 1b, in which participants were not offered a second response opportunity, we observed almost identical results. There was again highly significant improvement across the 10 sessions, as measured by accuracy, $F(1.38, 6.9) = 24.58$, $p = .001$, and information transfer, $F(1.41, 7.03) = 31.41$, $p < .001$. Accuracy improved from 22% to 46%, an average increase of 24%. Information transfer also increased, by an average of 0.83 bits, from 2.28 to 3.11 bits, which is equivalent to the perfect identification of approximately 8.66 stimuli. Even though the participants' average peak performance was greater in Experiment 1a than in 1b, this difference was not reliable according to an independent-samples *t* test ($p = .71$). Naturally, the statistical power of this test to identify between-experiment differences is limited, due to the small sample size. Nevertheless, we note that several participants in Experiment 1b showed larger practice effects than did some participants in Experiment 1a, making it seem unlikely that the two-response feedback procedure caused any large differences.

Discussion

Participants in both Experiments 1a and 1b demonstrated significant improvements in performance, suggesting that the two-response method was not responsible for the amount of learning observed. These experiments also confirm that Rouder et al.'s (2004) results were not due to unusual, virtuoso participants. In both experiments, performance improved by more than 20% (~0.8 bits) after about 6,000 practice trials, and three of the six participants in each experiment exceeded Miller's (1956) bounds of 7 ± 2 stimuli.

A possible explanation for the improvement in length judgment with practice might invoke the development of a relative (or "referent") judgment strategy rather than improvement in absolute identification processes themselves. That is, participants judging line lengths might be able to compare the lines with external magnitude cues, such as the edges of the computer monitor or the response buttons that appeared on screen. In Experiment 1, and in Rouder et al.'s (2004) design, these strategies were discouraged by jittering the absolute location of the stimuli on screen from trial to trial. Nevertheless, some small amount of relatively imprecise information might still be gained by comparisons with visible reference points, and it might be that this information alone supports improvement with practice. In Experiment 2, we investigated this explanation and also the idea that large effects of practice are possible only with large stimulus set sizes.

Experiment 2

Experiment 2 was conducted in a dark room. The edges of the monitor were obscured from view, and response buttons varied in size from trial to trial. Response buttons were never on-screen at the same time as stimuli. We also included a second condition, Experiment 2b, in which only half of the stimuli were presented, to determine whether the learning effect was due to the large amount of available information.

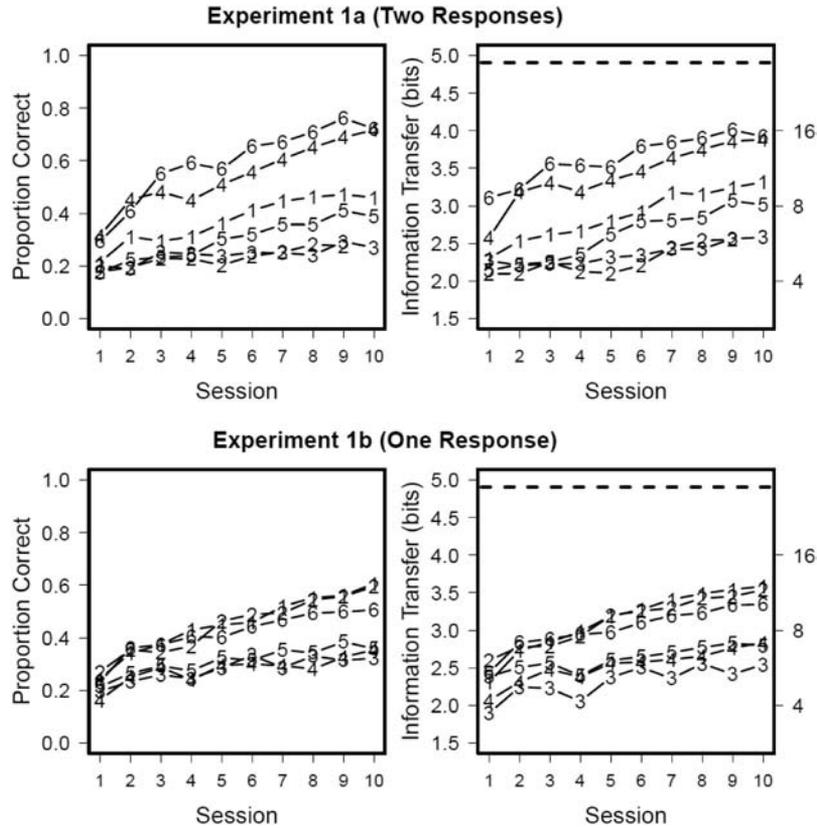


Figure 1. Proportion correct and information transfer as functions of session for Experiments 1a and 1b (30 lines varying in length). The right-hand axis on the information transfer graphs shows the equivalent number of stimuli that were perfectly identified (2^{bits}). The dashed lines indicate perfect performance: $\log_2(\text{number of stimuli})$.

Method

Participants. Ten participants took part in this experiment: five in Experiment 2a and five in Experiment 2b. They were reimbursed in a fashion similar to that for the participants in the first experiment.

Stimuli. Each stimulus consisted of a pair of white dots on a black background, horizontally separated by intervals that were of the same lengths as those of the lines in Experiment 1.

Procedure. There were two conditions defined by the number of stimuli: Experiment 2a had 30 stimuli, whereas Experiment 2b had only 15 stimuli. The stimuli in Experiment 2b were all of the odd-numbered stimuli from Experiment 2a, and so the pairwise stimulus separation was twice as large in Experiment 2b as in Experiment 2a. We could have kept stimulus separation equal in the two experiments (e.g., by presenting only Stimuli 1–15 in Experiment 2b), but that would instead have confounded stimulus range with set size. We acknowledge that both solutions to this problem (either confounding the stimulus range or separating the stimuli) are imperfect, but we chose the latter solution because performance is mostly unaffected by changes in stimulus for widely spaced stimulus sets (see e.g., Braida & Durlach, 1972; but also see Stewart et al., 2005, and Lacouture, 1997, for alternative findings).

The experiment was conducted in a dark room, where the only light was that emitted by the computer monitor (which was made as dark as possible). The edges of the computer monitor were obscured by black cardboard. To ensure that the response buttons could not be used as a cue for relative comparison with the size of the stimuli, two precautions were taken: The buttons were never present on-screen at the same time as the stimuli, and the size of the response buttons varied from trial to trial. That is, when a stimulus was presented, the buttons were removed from the screen until participants clicked a mouse button to indicate they were ready to respond, and then the stimulus was removed and the response buttons were displayed. Participants took part in 10 sessions, each about 1 hr in length. Each session consisted of six blocks of 90 trials, resulting in 180 presentations per stimulus in Experiment 2a and 360 presentations per stimulus for Experiment 2b.

Results

Performance increased significantly across the 10 sessions in both conditions. Participants in the 30-stimulus condition (Experiment 2a) increased their accuracy from 25% to 50%, an average improvement of 25%, $F(1.76, 7.04) = 29.74, p < .001$. Information transfer also increased, by 0.93 bits across the 10 sessions, from 2.44 bits to 3.36 bits, $F(1.54, 6.15) = 60.88, p < .001$, so the

average maximum performance was equivalent to perfect identification of approximately 10.3 stimuli.

Similarly, participants in the 15-stimulus condition also demonstrated highly significant improvements in both accuracy, $F(1.56, 6.24) = 22.25, p = .002$, and information transfer, $F(1.71, 6.86) = 19.43, p = .002$. Participants improved 33%: Average accuracy rose from 48% to 81%, and information transfer improved 1.08 bits, from 2.07 to 3.15. Average maximum information transmitted was equivalent to identification of 8.85 stimuli. Figure 2 provides a comparison of the individual participant results in Experiments 2a and 2b.

The average maximum amount of information transfer reported for the 30-stimulus condition in the current experiment ($M = 3.30$) was not reliably different from that found in Experiment 1a ($M = 3.25, p = .86$) and Experiment 1b ($M = 3.15, p = .50$).¹ This suggests that any external cues were not responsible for the learning effect in Experiment 1.

We also observed that, although one participant in the 15-stimulus condition reached almost perfect performance (94.3% accuracy), the average maximum information transmission for the 15-stimulus condition ($M = 3.11$) was not reliably different from those of the 30-stimulus conditions in Experiments 1a ($p = .64$) and 1b ($p = .85$) or Experiment 2a ($p = .38$). This suggests that maximum performance, in terms of information transmission, does not vary with set size.

Discussion

Participants in Experiment 2 demonstrated significant improvements in performance, and their information transmission limits and amounts of learning were similar to those of participants in Experiment 1. Once again, half of the participants demonstrated maximum information transfer rates that exceeded Miller's (1956) 7 ± 2 bounds. The similarity in results between Experiments 1 and 2 suggests that external cues were not responsible for the learning effect and that the amount of available information does not determine the extent of learning, at least as long as performance is below ceiling.

Experiment 3

So far, substantial learning in absolute identification has been demonstrated using only line lengths, with null (or small) effects observed tones varying in intensity or frequency. Unfortunately, this difference between stimulus modality has always been confounded with a procedural change: Tones were made available to participants for only a short period of time (typically 1 s), whereas lines were made available for as long as participants wished. There is some evidence to suggest that stimulus presentation time can influence performance. For example, Miller (1956) cited unpublished research by Irwin Pollack that found significantly smaller information transmission for lines varying in length when presented for short periods of time (2.6 bits), compared with longer presentation times (3.0 bits). Ward and Lockhead (1971) also found lower information transfer for a presentation time of 8 ms (0.19 bits) compared with presentation for 200 ms (1.07 bits), although they simultaneously manipulated luminance. In an attempt to examine whether unlimited presentation time may have

encouraged the learning effect, in Experiment 3 we masked line stimuli after 1 s—in line with usual practice for auditory stimuli.

Method

Six participants took part in Experiment 3, following the same procedure as used in Experiment 2a, with the exception of presentation time. Stimuli were left on the computer monitor for only 1 s, after which they were covered by a mask consisting of white dots scattered randomly over a rectangle of dimensions 1,024 pixels \times 684 pixels. The white dots were equal in size and luminance to the white dots used to construct the line stimuli. Masks remained on-screen until participants responded.

Results

The results are similar to those in Experiments 1 and 2a, in which participants were given stimuli with an unlimited presentation time. Accuracy increased from 24% to 42%, $F(1.56, 7.79) = 18.2, p = .002$, and information transfer increased from 2.39 bits to 3.06 bits, $F(1.54, 7.7) = 22.27, p < .001$. The average maximum performance (3.06 bits) was equivalent to the perfect identification of 8.32 stimuli (see Figure 3). Although average learning was slightly smaller, one participant still learned to identify more than Miller's (1956) upper limit of nine stimuli. In addition, the maximum amount of information transmitted with masked stimuli was only about 5% smaller than the average amount for Experiments 1 and 2, and this difference was not statistically reliable ($p = .30$). The similar pattern in results for the current experiment suggests that long presentation times were not required for the learning effect.

Discussion

Even with limited stimulus presentation times there was significant improvement in performance with practice, and these results were comparable to those in earlier experiments with unlimited viewing time. The slightly lower performance reached with short presentation times was not significantly different from that in previous experiments. The direction of the effect, however, suggests that limited presentation time, or perhaps the addition of the mask, may have limited the amount that participants could improve via practice, even if our sample sizes provided insufficient statistical power to detect a reliable difference. Most important, however, participants did still manage to significantly improve their performance, and the amount of improvement was not much smaller than that in Experiments 1 and 2.

Experiment 4

Experiments 1–3 established that learning was not due to the more unusual aspects of Rouder et al.'s (2004) methods or to external cues and that it was not much attenuated by a limitation on stimulus presentation time. In Experiment 4 we tested whether the

¹ Estimates of transmitted information are inflated by small sample sizes (see Norwich, Wong, & Sagi, 1998). For this reason, for comparisons between experiments we always calculated information transfer using data divided into fairly long (540-trial) segments.

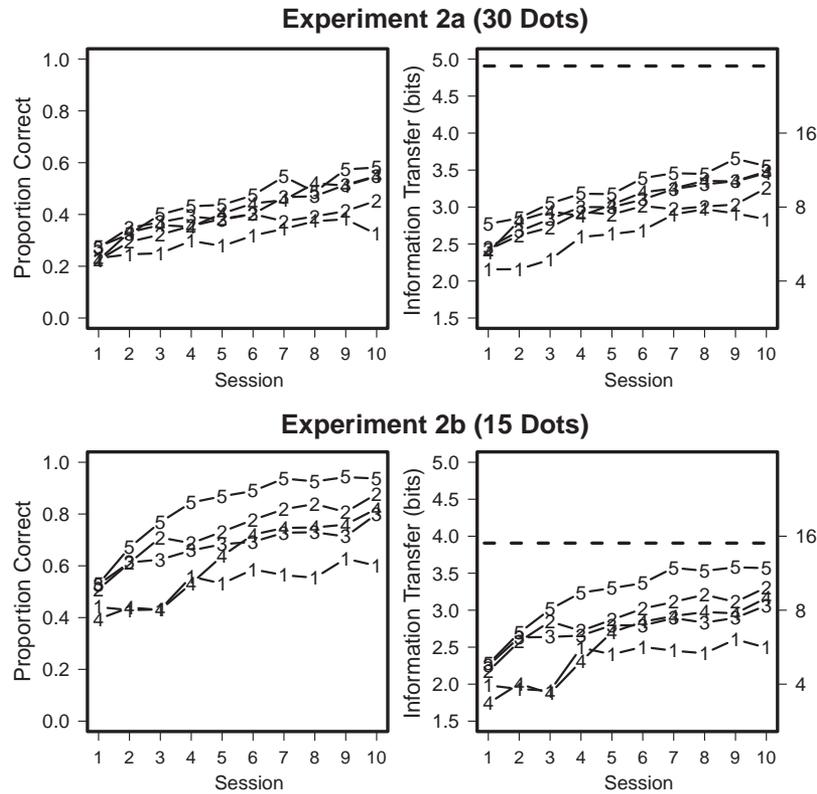


Figure 2. Proportion correct and information transfer as functions of session for Experiments 2a (30 stimuli) and Experiment 2b (15 stimuli), using dots varying in separation size. The right-hand axis on the information transfer graphs shows the equivalent number of stimuli that were perfectly identified (or 2^{bits}). The dashed lines indicate the maximum amount of information transfer possible: $\log_2(\text{number of stimuli})$.

strong practice effects we have observed are specific to visual lengths: lines varying in length or dots varying in separation. In this experiment we investigated whether learning is possible with lines varying in angle of inclination.

Method

The methods were identical to those in Experiment 3 except that the stimuli were 30 lines whose angle of inclination varied from

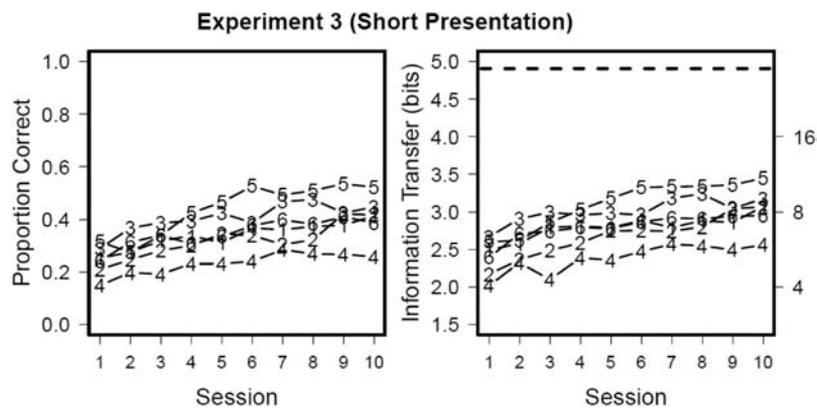


Figure 3. Proportion correct and information transfer as functions of session for Experiment 3 (dots varying in separation size). The right-hand axis on the information transfer graphs shows the equivalent number of stimuli that were perfectly identified (or 2^{bits}). The dashed lines indicate the maximum amount of information transfer possible: $\log_2(\text{number of stimuli})$.

1.5° to 89.5° in increments of 3°. The lines were 12-pixel \times 210-pixel rectangles, and they were blurred by applying Gaussian kernel with a 7-pixel standard deviation (to prevent the use of pixel aliasing as a cue for angle). Stimuli were white on a black background and were positioned within a square 300 pixels \times 300 pixels in size. To help prevent the use of both horizontal and vertical cues for angle judgments, we rotated lines around a central pivot point and randomly varied the screen position of that pivot point from trial to trial within a 22-pixel \times 22-pixel region. Each stimulus was presented for 1 s. If no response was made within 1 s, a mask was displayed and remained on-screen until participants made their response. Masks were 1,024-pixel \times 1,024-pixel squares containing a series of randomly positioned and randomly oriented lines of the same sort as the stimuli.

Results

Results were similar to those in the prior experiments. Figure 4 shows that learning was highly significant across the 10 sessions: accuracy, $F(2.05, 10.3) = 23.6$, $p < .001$; information transfer, $F(2.17, 10.8) = 26.37$, $p < .001$. Average accuracy improved by 22%, rising from an initial value of 24% to 46%. Average information transfer also improved, rising 0.81 bits, from 2.37 bits to 3.18 bits, which made average maximum performance equivalent to the perfect identification of about 9.05 stimuli. Three of the six participants exceeded Miller's (1956) 7 ± 2 limit after 10 sessions of practice. Concerning initial performance, performance improvement, and maximum performance, none were significantly different from those in Experiment 1a ($ps = .93, .68$, and $.78$, respectively) or Experiment 3, in which presentation time was identical ($ps = .87, .23$, and $=.37$, respectively).

Discussion

Participants in Experiment 4 demonstrated significant improvement in performance across the 10 sessions, similar to that observed in the previous experiments. This result suggests that the learning effect may generalize to visual stimuli other than line

length. In Experiment 5 we further explored whether learning occurs with other stimulus types.

Experiment 5

Clearly people are able to substantially improve their performance in an absolute identification task when given significant practice, and we have shown that this learning is not specific to distance or length judgments. However, so far our investigation has been limited to visual stimuli only. Miller (1956) noted that visual modalities led to slightly greater information transmission (hence the "plus or minus" in his 7 ± 2). More recent research has also suggested differences—Lacouture and Lacerte (1997) found better performance for lines varying in length than tones varying in intensity. This is particularly interesting here, because most previous studies showing no effect of practice used tones varying in intensity. In Experiment 5, we compared the effect of practice using tones varying in intensity and lines varying in length in order to determine whether it is modality that differentiates our (and Rouder et al.'s, 2004) findings from those of others.

Method

Participants. Six participants took part in Experiment 5a, and a different six took part in Experiment 5b.

Procedure. The procedure was identical to that in Experiment 1a except that stimuli were either 16 lines varying in length (Experiment 5a) or 16 tones varying in intensity (Experiment 5b). Though we aimed to replicate our earlier experiments exactly, we found that we were limited to the use of just 16 (rather than 30) tones. This limit was identified through pilot testing with naïve participants. Those tests showed that participants were able to make perfectly accurate discrimination judgments (lower/higher) between sequentially presented stimuli separated by a 1-s pause when the stimulus difference was 3 dB. This stimulus separation implied that range restrictions imposed by ethical considerations and the audio equipment itself limited us to 16 tones in total. We therefore also ran Experiment 5a using 16 lines (see Table 1 for

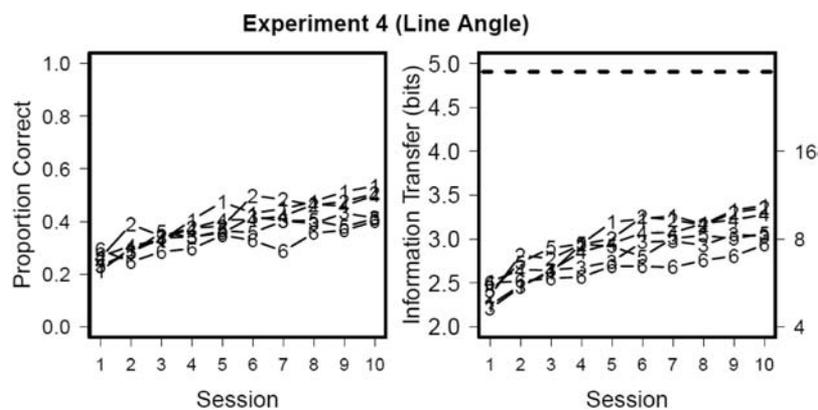


Figure 4. Proportion correct and information transfer as functions of session for Experiment 4 (lines' angle of inclination). The right-hand axis on the information transfer graphs shows the equivalent number of stimuli that were perfectly identified (or 2^{bits}). The dashed lines indicate the maximum amount of information possible: $\log_2(\text{number of stimuli})$.

line lengths in pixels) for ease of comparison with results of Experiment 5b. The 16 auditory stimuli were pure 1000-Hz tones, ranging from 61 dB to 106 dB, in 3-dB increments. Loudness was measured using a Brüel and Kjaer artificial ear (Model 4152) and sound level meter (Brüel and Kjaer, Model 2260), equipped with a condenser microphone (Brüel and Kjaer, Model 4144). Tones were played for 1 s each and were presented via Sony headphones (Model DR-220). For each of the 10 sessions, participants in the lines condition completed seven blocks of 80 trials, and those in the tones condition completed seven blocks of 90 trials.

Results

Participants in the 16-line condition performed similarly to participants in previously reported line experiments: Average accuracy significantly increased across the 10 sessions, from 49% to 78%, $F(1.82, 9.11) = 23.43, p < .001$, and average information transmission increased significantly from 2.34 bits to 3.15 bits, $F(1.94, 9.7) = 20.92, p < .001$. Average maximum performance was equivalent to identification of 8.86 stimuli, and two of the six participants exceeded Miller's (1956) 7 ± 2 limit. The maximum information limit reached in the 16-line condition was not significantly different from that in Experiment 1a ($p = .58$) or from the results in Experiment 2b with a similar set size ($p = .95$).

Participants in the tone intensity condition, on the other hand, failed to exhibit the substantial learning found in all other experiments (see Figure 5). Participants given 16 tones of varying intensity had a lower average initial accuracy (31%) and improved on average by only 12%. Similarly, information transfer increased on average by only approximately 0.46 bits, from 1.49 bits to 1.95 bits, meaning that maximum performance was equivalent to the perfect identification of only 3.86 stimuli, and no participant exceeded Miller's (1956) 7 ± 2 limit. In fact, all participants identified less than five stimuli perfectly correctly. However, the small effect of practice was statistically reliable: accuracy, $F(2.91, 14.2) = 4.8, p = .02$; information transfer, $F(2.08, 10.2) = 4.9, p = .03$.

Although the improvement for both modalities was reliable, loudness in Experiment 5b showed a significantly lower information transfer limit than did lines in Experiment 5a ($M_{\text{lines}} = 1.96, M_{\text{tones}} = 3.09, t(9.79) = 9.51, p < .001$). We also observed reliably smaller maximum information transmission, $t(9.47) = 3.17, p = .01$, for tones ($M_{\text{diff}} = 0.42$ bits) than for lines ($M_{\text{diff}} = 0.74$ bits). The findings for loudness were consistent with previous findings of a low channel limit (see e.g., Garner, 1953; Miller, 1956; Pollack, 1952) and little improvement with substantial practice (see e.g., Weber et al., 1977). It is also interesting that, in contrast

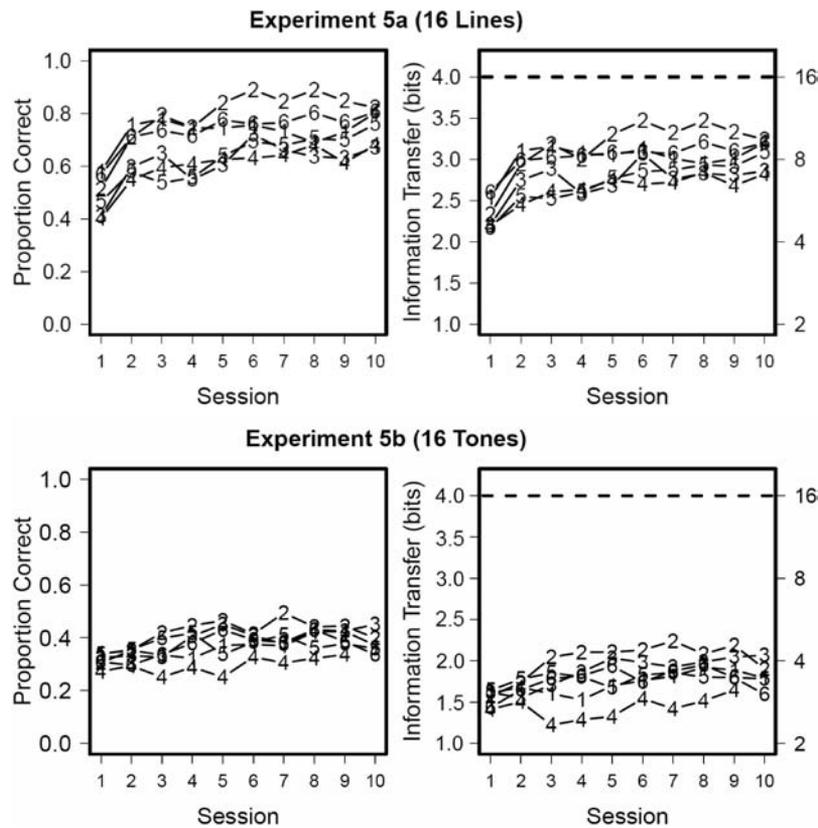


Figure 5. Proportion correct and information transfer as functions of session for Experiments 5a and 5b (lines varying in length and tones varying in loudness, respectively). The right-hand axis on the information transfer graphs shows the equivalent number of stimuli that were perfectly identified (or 2^{bits}). The dashed lines indicate the maximum amount of information transmission: $\log_2(\text{number of stimuli})$.

to the slow increase in performance for tones varying in intensity, there was a much faster increase in performance for line length (Experiment 5a). This was particularly noticeable between Sessions 1 and 2, where participants in Experiment 5a improved their performance significantly more ($M = .16$) compared with the corresponding difference between Session 1 and 2 in Experiment 5b ($M = .02$), $t(6.8) = 8.85$, $p < .001$. This suggests that participants in Experiment 5a (line stimuli) learned quickly to some upper limit, unlike participants in Experiment 5b.

To better understand the difference between learning with lines and with tone intensities, we further examined accuracy for each stimulus type. Figure 6 plots the proportion of correct identifications against ordinal stimulus magnitude, separately for the two stimulus continua and separately for data from the beginning and end of practice. When practicing with line lengths (Experiment 5a), there was general improvement for all stimuli across the range, except where limited by ceiling effects for the smallest and largest lines. Although not shown here, corresponding plots for all other experiments showed the same pattern as in Experiment 5a. However, for tone intensities, there was no reliable improvement for tones in the middle of the range (Stimuli 5–11). This suggests that the limited amount of learning we observed for tone intensities was restricted to tones near the ends of the range.

Discussion

Both the participants who practiced with 16 lines varying in length and those who practiced with 16 tones varying in intensity

demonstrated significant improvements in performance. Even though the improvement for both experiments was statistically reliable, participants who practiced with tone intensities showed a much smaller learning effect and a significantly lower information transfer limit than did those who practiced with lines. Participants practicing with tone intensities, in contrast to participants who practiced with other continua, also failed to improve their performance consistently across the stimulus range (see Figure 6).

Experiments 1–5 together suggest an interesting possibility—that the amount of improvement through learning is closely related to the initial level of performance, prior to practice. For example, initial accuracy with tones of varying intensity (Experiment 5b) was poorer than for any other stimulus continuum and so was the amount of improvement with practice. Conversely, accuracy with lines of varying length was initially high and so was the amount of improvement with practice. In Experiments 6 and 7, we explored the relationship between initial performance and learning and further investigated the generality of the learning effect across different stimulus dimensions, using tone frequency.

Experiment 6

Experiment 6 uses tones of varying frequency. Other research (see e.g., Garner, 1953; Pollack, 1952; Stewart et al., 2005) has shown that prepractice performance with tone frequency is similar to, but slightly better than, that for tone intensity. From this, we hypothesized that the amount of improvement from practice would be a little more than that observed for tones of varying intensity but still less than that observed for lines of different length.

Method

Participants. Six new participants took part in Experiment 6.

Stimuli. Stimuli were 36 tones varying in frequency. The range of frequencies (see Table 2) mimicked piano key frequencies, ranging from A3 to G#5 (220.0 Hz to 1661.2 Hz). Tones were pure sine waves, generated using Matlab R2008b, and were presented via headphones at a constant sound pressure, corresponding to 75 dB at 1000 Hz.

Procedure. On each trial, a fixation cross appeared for 500 ms, before the tone was played through the headphones for 1 s. Participants were free to respond either during or after playback. Feedback was as in Experiment 1a: Participants were given two response opportunities. Buttons were available on-screen in three horizontal rows of 12, and participants responded using the mouse. The buttons had not only the numerical label normally associated with absolute identification (i.e., 1, . . . , 36) but also the corresponding piano key note (i.e., A3, . . . , G#5). Three of the six participants had some musical training; the other three had none at all.

In eight of the 10 sessions, participants practiced for six blocks of 108 trials each. In the first and last session however, participants completed only four blocks of 108 trials. Fewer experimental trials were completed in this first and last session because participants also completed a brief pairwise discrimination task. This task consisted of two blocks of 72 trials, during which participants were asked to discriminate between adjacent stimuli in the set. Adjacent tones were presented sequentially—the first tone was played for 1 s, followed by 500 ms of silence, and then either the

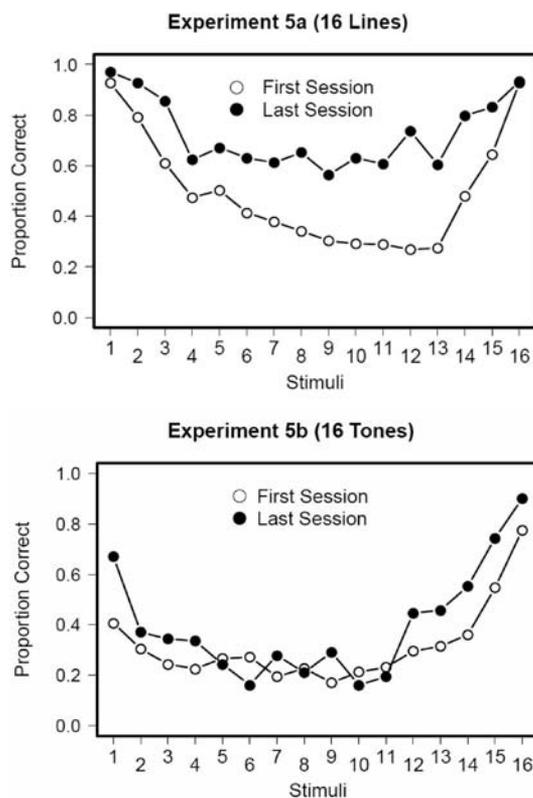


Figure 6. The proportion of correct identifications plotted against ordinal stimulus magnitude for Experiments 5a (lines) and 5b (tone intensity) for both the first and the last 540 trials.

Table 2
Range of Frequencies Used in Experiments 6 and 7

					Frequencies						
220.0	233.1	246.9	261.6	277.2	293.7	311.1	329.6	349.2	370.0	392.0	415.3
440.0	466.2	493.9	523.3	554.4	587.3	622.3	659.3	698.5	740.0	784.0	830.6
880.0	932.3	987.8	1046.5	1108.7	1174.7	1244.5	1318.5	1396.9	1480.0	1568.0	1661.2

Note. Frequencies correspond to musical notes on a keyboard from A3 to G#5.

next higher or next lower frequency tone in the stimulus set was played for 1 s. The participant was then asked to indicate which of the two tones was higher. This pairwise discrimination task simply confirmed that all participants were perfectly able to discriminate between adjacent stimuli, both before and after practice.

Results

The added pairwise discriminability task meant that the number of trials in Sessions 1 and 10 of Experiment 6 was not equal to those in other sessions. Because information transfer is sensitive to sample size, “pseudosessions” of 540 trials each were used for analysis. Two participants were unable to complete all six blocks of the experiment within the allotted time frame in each session and hence completed fewer trials (4,104 and 3,996 trials for each participant, respectively, equivalent to seven pseudosessions of 540 trials) than did other participants (from 5,940 to 6,264 trials, or 11 pseudosessions). The lines for individual participants in Figure 7 reflect this imbalance in trial numbers.

Those participants who completed 11 pseudosessions demonstrated a reliable improvement in information transfer, from 2.21 bits to 2.59 bits (6.02 stimuli) and in accuracy, from 19% to 30%: respectively, $F(2.47, 7.42) = 7.18, p = .02$, and $F(1.38, 4.13) = 9.82, p = .03$. One participant was quite different from the others, and only this person exceeded Miller’s (1956) bounds of 7 ± 2 stimuli, identifying the equivalent of 14.4 stimuli. This exceptional participant was one of the three participants in this experiment who had several years of musical training (the other

two such participants performed the same as did the three untrained participants).

Discussion

As expected, the level of performance in the initial session was a little better than that for tone intensity (Experiment 5b) but lower than for all our experiments with visual stimuli. In line with our hypothesis, the amount of improvement due to practice was also greater than for tone intensity but less than that observed for comparable visual experiments. One remarkable participant showed a large improvement with learning, even relative to the experiments with visual stimuli. The exceptional participant was the one with the most musical experience and also the participant who began musical training at the youngest age. These facts agree with findings from the absolute pitch literature, suggesting that early and lengthy musical training encourage the development of absolute pitch (see e.g., Takeuchi & Hulse, 1993). Newer research has also suggested that absolute pitch ability exists at a baseline rate in the general population of people with little musical training (see e.g., Ross, Olson, & Gore, 2003). We are examining the relationship between absolute pitch and absolute identification, as well as the effect of practice on each, in experiments currently under way in our laboratory.

Experiment 7

We noted that better initial performance was correlated with greater improvement through practice. However, in Experiments

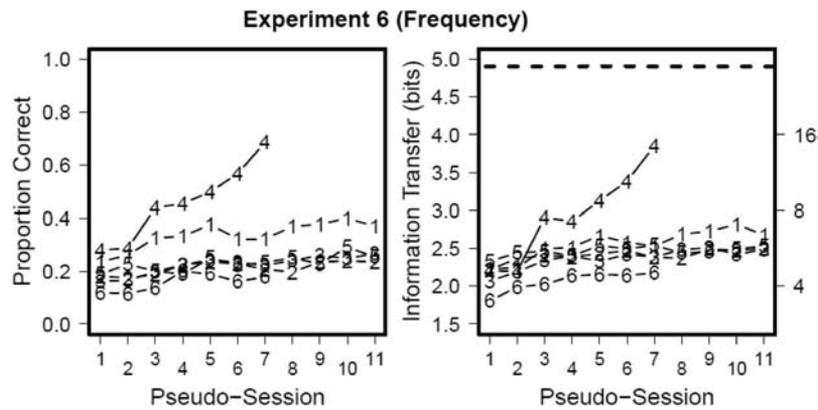


Figure 7. Proportion correct and information transfer for Experiment 6 (tones varying in frequency). The right-hand axis on the information transfer graph shows the equivalent number of stimuli that were perfectly identified (or 2^{bits}). Each pseudosession is equivalent to 540 trials. Two participants completed fewer trials than did other participants and hence have data for only seven pseudosessions. The dashed line indicates the maximum possible information transmission: $\log_2(\text{number of stimuli})$.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

1–6, this correlation was observed across different stimulus manipulations. That is, some kinds of stimuli support better initial performance than do others, and these also tend to support greater learning effects. In Experiment 7, we decoupled initial performance level from any stimulus manipulations, by instead manipulating participants' motivation.

Method

Six new participants completed Experiment 7, which replicated Experiment 6 but with one methodological difference: Participant reimbursement was contingent on performance. Correct and incorrect responses were rewarded differently (see Table 3), but we provided a minimum reimbursement of \$150 for 10 sessions. The maximum reimbursement actually achieved by a participant was just over \$220.

Results

Similar to in Experiment 6, performance reliably increased across the 10 sessions for both accuracy and information transfer (see Figure 8): accuracy, $F(1.47, 7.35) = 10.85, p = .009$; information transfer, $F(1.63, 8.16) = 14.06, p = .003$. In line with our hypothesis that motivation may be manipulated by monetary incentive, the average amount of improvement was larger, $t(8.8) = 2.39, p = .04$, in Experiment 7 than in Experiment 6 (an average of 0.38 bits for 11-session participants in Experiment 6, compared with 0.64 bits in Experiment 7), but the difference in initial performance was nonsignificant (2.14 bits for 11-session participants in Experiment 6 compared with 2.15 bits in Experiment 7; $p = .87$).

A parametric test of the difference between the improvement seen in Experiments 6 and 7 is inappropriate, due to the exceptional participant in Experiment 6. Consequently we used a non-parametric Wilcoxon test, which takes account of only ordinal (rank) information and so is not unduly distorted by the virtuoso participant. The results of this test supported the hypothesis that participants who received motivational reimbursements improved their performance more than did those who did not ($W = 29, p = .047$).

Discussion

Participants who received extra motivation through monetary reimbursement based on response accuracy showed almost twice as much learning as did those who received reimbursement independent of performance. This result is consistent with Rouder et al.'s (2004) suggestion that motivation is required for learning. Experiment 7 also has implications for the effect of stimulus modality on learning. If stimulus modality were the only deter-

mining factor for practice effects, we would have expected little difference between Experiments 6 and 7. Instead, when participants were suitably motivated we observed that they improved by slightly less than did those in the visual modality experiments but much more than in the two other auditory experiments.

Although the amount of improvement observed in Experiment 7 compared with that in Experiment 6 emphasizes the importance of motivation, there was no significant difference between initial (first session) performance levels in the two experiments. This makes it difficult to directly evaluate the hypothesis that initial performance predicts overall improvement. Future research could produce a more direct test of the hypothesis by experimentally manipulating the initial performance level. However, we note that our initial attempts at such experiments have proven unsatisfactory, because almost all manipulations that influence initial performance level involve manipulations of the stimuli, thus confounding the critical hypothesis with other hypotheses regarding stimulus-driven effects.

Regardless of the motivational manipulations in Experiment 7, however, no participant came close to performing as well as did the one exceptional participant in Experiment 6 (who performed better than did the majority of participants across all experiments). This finding speaks to the strength of individual differences in absolute identification performance, both in initial performance and amount of learning. We analyze these individual differences across experiments in the next section.

Summary of Results

Table 4 contains a summary of information transmission results from the 10 conditions in our seven experiments. Through these experiments we have shown that the learning observed for lines of varying length in Rouder et al. (2004) was not due to virtuoso participants or atypical methodological aspects of their design: Experiment 1 showed that learning was not due to the two-response method, and Experiment 2 showed that learning was not due to external visual cues. Experiment 3 showed that the extended stimulus presentation time associated with lines compared with auditory stimuli was not required for learning. In Experiment 4 we showed that substantial learning also occurred for another visual stimulus—lines of varying inclination. Experiment 5 showed small practice effects for tones of varying intensity. Consistent with the conventional wisdom about absolute identification (see e.g., Shiffrin & Nosofsky, 1994), participants reached a low information transmission limit after a few sessions. Experiments 6 and 7 demonstrated that learning was possible for another auditory continuum, particularly when participants were well motivated. Participants practicing with tones of varying frequency were able to learn much more than those with tones of varying intensity, but not as much as those who practiced with most of our visual stimuli.

Rouder et al.'s (2004) results were surprising because they violated two truisms of absolute identification: that practice has little effect on performance and that there is a severe limitation in performance, equivalent to 7 ± 2 stimuli. Our results confirm and generalize Rouder et al.'s observation that practice can have a substantial effect on performance. However, the last column in Table 4 shows that on average, participants did not greatly exceed Miller's (1956) limit of nine stimuli after 10 hr of practice. Indeed, the equivalent number of stimuli perfectly identified after practice,

Table 3
Rate of Reimbursement Used in Experiment 7

Response attempt	Response accuracy		
	Correct	One off	Two off
First	\$0.05	\$0.03	\$0.02
Second	\$0.01	\$0.00	\$0.00

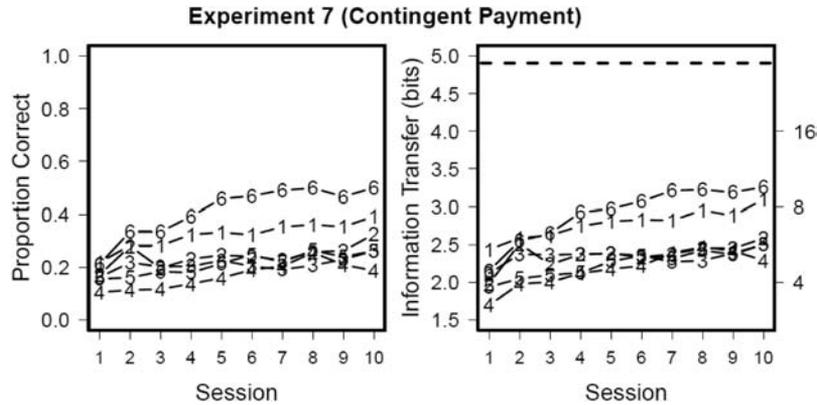


Figure 8. Proportion correct and information transfer for Experiment 7 (tones varying in frequency, using contingent payment methods). The right-hand axis on the information transfer graph shows the equivalent number of stimuli that were perfectly identified (or 2^{bits}). The dashed line indicates the maximum amount of information possible: $\log_2(\text{number of stimuli})$.

averaged across visual stimuli, for which performance was best, was 9.88 stimuli, not much above Miller’s upper limit.

Individual participants, however, tell a different story. Of the 58 participants who took part in all experiments, 22 exceeded Miller’s (1956) limit. Indeed, two participants (in Experiments 1a and 6) reached a maximum rate of information transfer over 4 bits in their last session (16.1 and 17.5 stimuli, respectively). These results are reminiscent of participant RM in Rouder et al. (2004), who was able to perfectly identify approximately 20 stimuli. Given that Rouder et al. looked at the effect of practice for only three participants, it is possible that their participants are best thought of as equivalent to the better performers in our experiments. Indeed, given that two of their three performers were authors, we expect their results are also consistent with our findings regarding improvements due to increased motivation. It seems, therefore, that Miller’s magical number 7 ± 2 may be best interpreted as not being too far wrong for the average participant, even if this is not true for some individuals.

General Discussion

The deeper question our work has provoked is, what produces differences in the ability to increase capacity in absolute identifi-

cation? Table 4 shows that the experiments in which there was a large amount of improvement with practice were also the same experiments in which performance during the first practice session was high. Figure 9 shows that this result extends, at least approximately, to the individual participant level. That is, participants who performed well in their first practice session—no matter which experiment they were in—also tended to be those who showed large learning effects. Although far from perfect, the correlation between initial performance and improvement was substantial both for experiments with larger set sizes (30 or 32 stimuli), $r(39) = .609$, and for those with smaller set sizes (15–16 stimuli), $r(15) = .653$, both $ps < .001$. Participants’ levels of initial performance were highly correlated with the stimulus modality used for their experiment, $r(54) = .94$, $p < .001$, but a partial correlation confirmed that individual differences in initial performance still explained unique variance in the amount of improvement with practice, even after removing the effects of stimulus modality, $r(53) = .40$, $p < .01$.

The correlational analyses suggest that two important determinants of learning are (a) stimulus continuum (modality) and (b) individual differences between participants, at least partly caused by motivation, and that the effects of both factors are well de-

Table 4
Summary of Results for All Experiments

Experiment	Stimulus continuum	Set size	Average information (bits)			
			Session 1	Minimum	Improvement	Maximum
1a	Lines (length)	30	2.38	2.37	0.88	3.25
1b	Lines (length)	30	2.28	2.28	0.87	3.15
2a	Dots (separation)	30	2.45	2.45	0.85	3.30
2b	Dots (separation)	15	2.08	2.08	1.03	3.11
3	Dots (separation)	30	2.39	2.39	0.64	3.03
4	Lines (angle)	30	2.37	2.37	0.80	3.17
5a	Lines (length)	16	2.35	2.35	0.74	3.09
5b	Tones (intensity)	16	1.56	1.53	0.42	1.96
6 ^a	Tones (frequency)	36	2.14 (2.21)	2.14 (2.21)	0.59 (0.38)	2.73 (2.59)
7	Tones (frequency)	36	2.15	2.15	0.64	2.78

Note. All results are calculated on the basis of pseudosession, or every 540 trials, for ease of comparison.

^a For Experiment 6 the averages in parentheses represent those participants who completed 11 pseudosessions.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

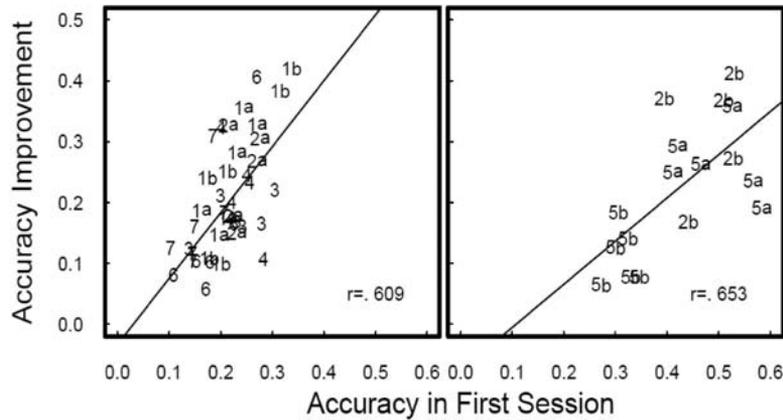


Figure 9. Improvement as a function of accuracy in the first 540 trials for experiments with larger set sizes (30 or 32 stimuli; left panel) and smaller set sizes (15 or 16 stimuli; right panel). Each point represents a single participant from a single experiment. The number denoting each participant on the graph is the experiment in which he or she took part.

scribed by performance during the first hour of experimental trials. Rast and Zimprich (2009) also found both strong individual differences and a positive correlation between participants' initial performance and learning rate in paired associate learning. This task bears some resemblance to absolute identification, in which participants must learn stimulus-label associations (Siegel & Siegel, 1972). Note that the observed positive correlation between initial performance and learning need not have occurred. For example, a naïve expectation might have been a negative correlation, because higher initial performance leaves less room for improvement. Indeed, such a result seems assured in extreme cases in which ceiling effects arise, such as for participants with almost perfect initial performance levels.

Our results do not uniquely identify the mechanism through which increased initial performance might be associated with greater overall improvement. However, several mechanisms seem likely candidates. For example, an exemplar model may naturally account for such improvement if information about the magnitude of a stimulus is stored only when the response is correct. A second possible explanation for the differences between experiments is that they depend on the pairwise discriminability of the stimulus sets, which might similarly vary between participants. The Weber fractions for length and loudness are approximately 2% and 4.8%², respectively (Laming, 1986; Teghtsoonian, 1971), suggesting that people are less sensitive to changes in tones of varying loudness than lines of varying length. Such explanations seem implausible, however, because in all of our experiments, stimulus separation was well above the Weber fraction, and research has shown that increasing separation between stimuli either has no effect at all (see e.g., Gravetter & Lockhead, 1973; Pollack, 1951) or results in a quite small improvement in performance (see e.g., Lacouture, 1997; Stewart et al., 2005).

Theoretical Implications

Recent years have seen the development of several comprehensive models for performance in absolute identification (see e.g., Brown, Marley, Donkin, & Heathcote, 2008; Kent &

Lamberts, 2005; Petrov & Anderson, 2005; Stewart et al., 2005). Our findings present severe challenges for these theories on several fronts, challenges that may require substantial redevelopment of the models. Such development is beyond the scope of this article, and so we limit ourselves to delineating the problem and providing an example of the direction that model development could take.

All theories of absolute identification respect the received wisdom in the field. No modern theories include any mechanism by which sustained practice can improve performance, and all theories take pains to treat all stimulus continua identically, as long as pairwise discrimination is perfect. Both of these assumptions are challenged by our results and those reported by Rouder et al. (2004)—namely, that theories must predict learning with practice and that this learning should be different for different stimulus continua. The third major challenge for theoretical accounts is to accommodate the correlation we observed between initial performance and the amount of improvement with practice. It is not yet obvious to us how to develop a theory for absolute identification that accommodates our results in a natural way. However, as a proof of concept, we illustrate that it is possible to build learning effects into the selective attention, mapping, and ballistic accumulation (SAMBA) model for absolute identification (Brown et al., 2008). Similar illustrations are likely able to be constructed for other models.

Most theoretical accounts of performance in absolute identification agree that incorrect responses arise from two separate sources—systematic biases and capacity limitations—and it is reasonable to posit that learning may improve performance by acting on either source. Contrast is one important systematic effect for the former type; contrast is where decisions are biased

² Laming (1986) observed that the Weber fraction for pure tones improves as intensity increases according to the function $0.23A^{-0.14}$, where A is the amplitude. The magnitudes used in the current study (61 dB–106 dB) are sufficiently large to make Laming's function well approximated by the constant Weber fraction reported.

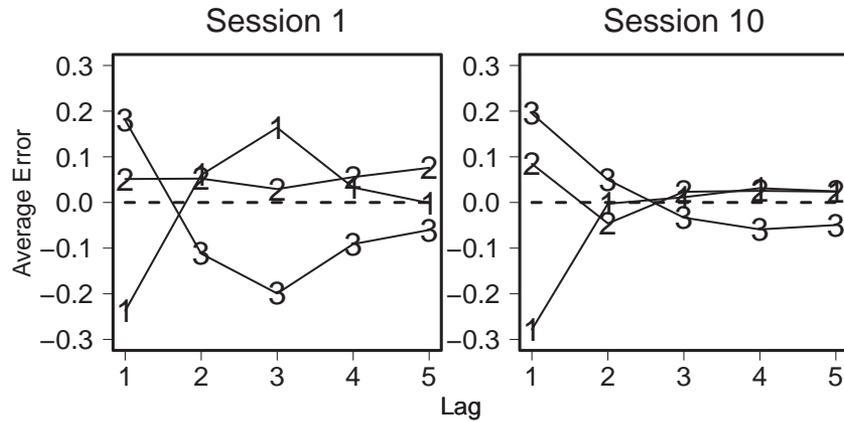


Figure 10. Impulse plots for data from Session 1 (left panel) and Session 10 (right panel) in Experiment 1a. The different lines represent the magnitude of the stimulus presented one to five trials previously: Line 1 = Stimuli 1–10, Line 2 = Stimuli 11–20, and Line 3 = Stimuli 21–30. The x-axis (lag) shows the number of trials since the occurrence of the stimulus used to condition the three lines. The dotted line provides a reference point showing where perfectly correct responses would lie.

away from stimuli observed a few trials earlier (e.g., if one observed a large-magnitude stimulus two or three trials previously, the current decision is likely to be biased toward smaller responses). Like all systematic biases, contrast reduces accuracy, but Treisman and Williams (1984) showed how contrast can be viewed as an adaptive mechanism that helps the observer track changes in a nonstationary stimulus environment. For example, SAMBA (Brown et al., 2008) attributes contrast effects to the redirection of selective attention toward recently seen magnitudes. This improves performance in a changing environment by keeping attention directed toward relevant stimulus magnitudes. Although this mechanism is adaptive in general, and particularly when the stimulus set is unfamiliar, our experiments employed a fixed set of stimulus magnitudes for thousands of trials, making tracking unnecessary. In this case, contrast impedes performance without any benefit, and so it would be rational to reduce contrast with practice.

Analysis of the data from Experiment 1a support this notion. Figure 10 illustrates sequential effects in the data from Experiment 1a, using an impulse plot (Ward & Lockhead, 1971). It shows average error as a function of the number of trials since stimulus presentation for data from the first and last sessions of Experiment 1, averaged over participants and over groups of 10 adjacent stimuli (i.e., Line 1 represents Stimuli 1–10, Line 2 represents Stimuli 11–20, and Line 3 represents Stimuli 21–30). The data from the first session of practice (left panel) show standard bias effects: assimilation of the responses toward the stimulus from the previous trial and contrast of responses away from stimuli from earlier trials. For example, when a small stimulus (Line 1) was shown on the previous trial (lag = 1), average errors were negative, meaning that responses tended to be smaller than the correct response (i.e., errors are biased toward the previously presented small stimulus). When the same small stimulus was presented a few trials previously (lag > 1), the data show contrast, where

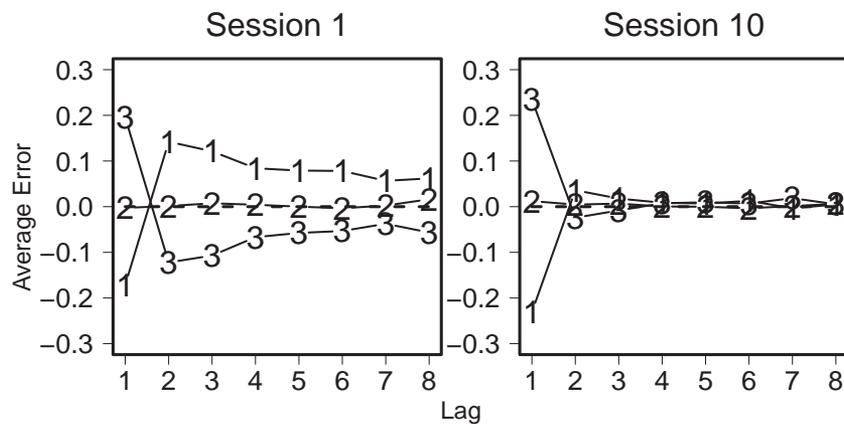


Figure 11. Impulse plots for SAMBA (selective attention, mapping, and ballistic accumulation) simulations for Session 1 (left panel) and Session 10 (right panel) with decreasing contrast. The dotted line provides a reference point showing where perfectly correct responses would lie.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

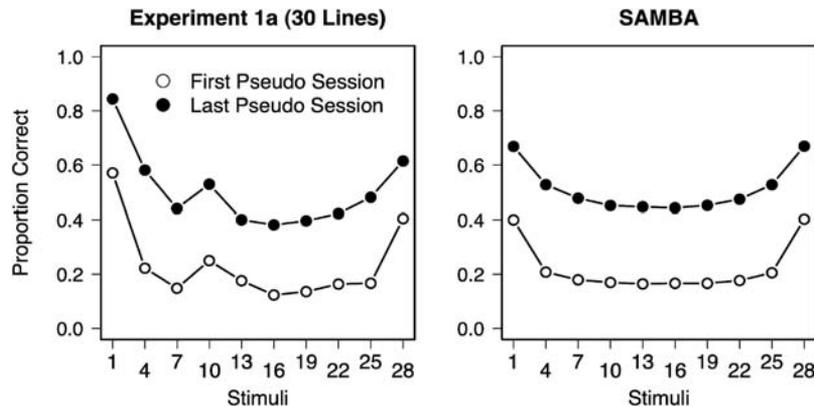


Figure 12. Response accuracy versus (rank) stimulus magnitude for Experiment 1a (left panel) and SAMBA (selective attention, mapping, and ballistic accumulation) predictions (right panel). Open and filled symbols correspond to data from the beginning and end of practice, respectively. The data are averaged over participants and over groups of three consecutive stimuli (e.g., the left-most point on each line represents average accuracy for Stimuli 1, 2, and 3).

errors tended to be too large when the stimulus presented two or more trials ago was small (i.e., biased away from the previously presented small stimulus). Data from the final session of practice (right panel) are unusual: The magnitude of the contrast effect decreased markedly with practice, whereas the assimilation effect did not change much.

Theoretical considerations and the data both suggest that one way to include learning effects in absolute identification is by reducing the magnitude of model parameters governing contrast without altering assimilation. This approach fits naturally with the SAMBA model because SAMBA attributes contrast effects to a selective attention process but assimilation effects to a more automatic, lower level inertia in the decision process. To simulate this process in SAMBA, we began by setting all parameters at values estimated by Brown et al. (2008; to fit a data set from Lacouture, 1997, parameters reported in Brown et al.'s Table 2). Then, to match the data from the first session of Experiment 1a we adjusted three parameters: We reduced the size of the assimilation parameter ($D = .035$), increased the size of the contrast parameter ($M = .25$), and adjusted the rehearsal capacity ($\lambda = .872$) to match the overall accuracy level of the data. The predicted impulse plot for the model using these parameters is shown in the left panel of Figure 11. To simulate the result of learning by reducing the contrast magnitude, we steadily reduced the contrast parameter to $M = 0$ over the course of learning, which removes almost all contrast effects from the model's predictions for the final session, as shown in the right panel of Figure 11.

When the effect of learning is modeled by the reduction of the contrast magnitude, the impulse plots predicted by SAMBA match the data quite well. However, this way of modeling learning fails to capture the large improvement in accuracy. To match the large accuracy gains made by participants, the model also needs to have its rehearsal capacity parameter changed with practice (from $\lambda = .875$ to $\lambda = 10$). This version of the model, with both contrast and rehearsal capacity influenced by practice, matches both the impulse plots and the accuracy data from Experiment 1a. The left panel of Figure 12 shows the proportion of correct responses in Experiment 1a as a function

of ordinal stimulus magnitude separately for the first and last sessions of practice, and the right panel shows the same calculations for the predictions of SAMBA given the aforementioned parameter values. Although the model parameters were not adjusted to accommodate all effects (such as the tendency in the empirical data for better performance with small than large line lengths), SAMBA accounts well for the effect of practice. As in the data, the model predicts a substantial increase in performance over practice, and this increase is approximately equal in magnitude across the range of stimuli. SAMBA also predicts an increasing U-shape in this plot with practice, and the data appear to confirm this prediction.

Similar accounts could be implemented in other comprehensive models of absolute identification, because they all include separate contrast and capacity parameters that can be manipulated as was just done. This approach advances theoretical understanding because it delimits the mechanisms by which practice improves performance, greatly constraining model development. However, there are three important questions that are left unaddressed:

1. By what mechanism(s) are rehearsal capacity and contrast magnitude changed by practice?
2. Why are there differences in the effect of practice when using different stimulus continua?
3. Why should prepractice performance correlate strongly with the amount of improvement from practice?

Conclusions

Rouder et al. (2004) demonstrated that practice dramatically improved performance in absolute identification. We have shown that this effect generalizes across most participants and many different procedural manipulations. We also found reliable effects of some stimulus manipulations, a surprising correlation between initial performance and the gains from practice, and a dissociation

between the effects of practice on assimilation and contrast magnitude. We showed that the fundamental result (improved accuracy with practice), as well as the dissociation, can be accommodated quite naturally within an existing comprehensive theory of absolute identification. The remaining findings stand as a challenge for the field: to develop a theory that naturally predicts improved performance and decreased contrast with practice, as well as provides a link between initial and final performances. A theory that provides such a link might then also explain the differences observed between stimulus continua, because many of the differences in amount of learning between continua were captured by differences between initial performances on those continua.

References

- Braida, L. D., & Durlach, N. I. (1972). Intensity perception. II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America*, *51*, 483–502. doi:10.1121/1.1912868
- Brown, S. D., Marley, A. A. J., Donkin, C., & Heathcote, A. (2008). An integrated model of choices and response times in absolute identification. *Psychological Review*, *115*, 396–425. doi:10.1037/003-95X.115.2.396
- Garner, W. R. (1953). An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology*, *46*, 373–380. doi:10.1037/h0063212
- Garner, W. R., & Hake, H. W. (1951). The amount of information in absolute judgments. *Psychological Review*, *58*, 446–459. doi:10.1037/h0054482
- Gravetter, F., & Lockhead, G. R. (1973). Criterial range as a frame of reference for stimulus judgment. *Psychological Review*, *80*, 203–216. doi:10.1037/h0034281
- Hartman, E. B. (1954). The influence of practice and pitch distance between tones on the absolute identification of pitch. *American Journal of Psychology*, *67*, 1–14. doi:10.2307/1418067
- Kent, C., & Lamberts, K. (2005). An exemplar account of the bow and set-size effects in absolute identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 289–305. doi:10.1037/0278-7393.31.2.289
- Lacouture, Y. (1997). Bow, range, and sequential effects in absolute identification: A response-time analysis. *Psychological Review*, *60*, 121–133.
- Lacouture, Y., & Lacerte, D. (1997). Stimulus modality and stimulus-response compatibility in absolute identification. *Canadian Journal of Experimental Psychology*, *51*, 165–170. doi:10.1037/1196-1961.51.2.165
- Laming, D. (1986). *Sensory analysis*. London, England: Academic Press.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits in our capacity for processing information. *Psychological Review*, *63*, 81–97. doi:10.1037/h0043158
- Norwich, K. H., Wong, W., & Sagi, E. (1998). Range as a factor determining the information of loudness judgments: Overcoming small sample bias. *Canadian Journal of Experimental Psychology*, *52*, 63–70. doi:10.1037/h0087281
- Petrov, A. A., & Anderson, J. R. (2005). The dynamics of scaling: A memory-based anchor model of category rating and absolute identification. *Psychological Review*, *112*, 383–416. doi:10.1037/0033-295X.112.2.383
- Pollack, I. (1951). Sensitivity to differences in intensity between repeated bursts of noise. *Journal of the Acoustical Society of America*, *23*, 650–653. doi:10.1121/1.1906815
- Pollack, I. (1952). The information of elementary auditory displays. *Journal of the Acoustical Society of America*, *24*, 745–749. doi:10.1121/1.1906969
- Rast, P., & Zimprich, D. (2009). Individual differences and reliability of paired associates learning in younger and older adults. *Psychology and Aging*, *24*, 1001–1006. doi:10.1037/a0016138
- Ross, D. A., Olson, I. R., & Gore, J. C. (2003). Absolute pitch does not depend on early musical training. *Annals of the New York Academy of Sciences*, *999*, 522–526. doi:10.1196/annals.1284.065
- Rouder, J. N. (2001). Absolute identification with simple and complex stimuli. *Psychological Science*, *12*, 318–322. doi:10.1111/1467-9280.00358
- Rouder, J. N., Morey, R. D., Cowan, N., & Pfaltz, M. (2004). Learning in a unidimensional absolute identification task. *Psychonomic Bulletin & Review*, *11*, 938–944.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423.
- Shiffrin, R. M., & Nosofsky, R. M. (1994). Seven plus or minus two: A commentary on capacity limitations. *Psychological Review*, *101*, 357–361. doi:10.1037/0033-295X.101.2.357
- Siegel, J. A., & Siegel, W. (1972). Absolute judgment and paired-associate learning: Kissing cousins or identical twins? *Psychological Review*, *79*, 300–316. doi:10.1037/h0032945
- Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review*, *112*, 881–911. doi:10.1037/0033-295X.112.4.881
- Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin*, *113*, 345–361. doi:10.1037/0033-2909.113.2.345
- Teghtsoonian, R. (1971). On the exponents in Stevens' Law and the constant in Ekman's Law. *Psychological Review*, *78*, 71–80. doi:10.1037/h0030300
- Thorndike, E. L. (1932). *The fundamentals of learning*. New York, NY: Teachers College Press. doi:10.1037/10976-000
- Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, *91*, 68–111. doi:10.1037/0033-295X.91.1.68
- Ward, L. M., & Lockhead, G. R. (1971). Response system processes in absolute judgment. *Perception & Psychophysics*, *9*, 73–78.
- 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.

Received November 15, 2009

Revision received October 3, 2010

Accepted October 14, 2010 ■