Title
Revising the limits of learning in Absolute Identification

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 31(31)

ISSN
1069-7977

Authors
Brown, Scott
Dodds, Pennie
donkin, Christopher
et al.

Publication Date
2009

Peer reviewed

Revising the limits of learning in Absolute Identification

Pennie Dodds (Pennie.Dodds@newcastle.edu.au)
School of Psychology, University Drive
Callaghan, NSW, Australia

Chris Donkin (Chris.Donkin@newcastle.edu.au)
School of Psychology, University Drive
Callaghan, NSW, Australia

Scott Brown (Scott.Brown@newcastle.edu.au)
School of Psychology, University Drive
Callaghan, NSW, Australia

Andrew Heathcote (Andrew.Heathcote@newcastle.edu.au)
School of Psychology, University Drive
Callaghan, NSW, Australia

Abstract

Miller's (1956) review of a series of absolute identification (AI) experiments, as well as a multitude of subsequent absolute identification research, suggests a fundamental limit to human information processing capacity. This limit is thought to be highly resistant to practice, independent of stimulus modality, and has been universally accepted as a fundamental constraint on human information processing capacity. Generally it is expected that people improve their performance slightly in absolute identification tasks, but quickly reach an asymptote after which they fail to improve any more. Recently however, we have replicated an experiment that demonstrates significant improvement in AI performance with only moderate practice. We conclude that there are several factors that are essential to the ability to learn unidimensional AI stimuli. Motivation is essential for improvement in performance, as is an initial performance level that greatly exceeds what would be expected by chance – this also constrains the type of stimuli that can be learned. In addition, in contrast to Miller's conclusion that the asymptote in performance is independent of set size, we suggest that indeed set size does affect the asymptote in performance, namely that a larger set size (around n=30), allows a higher asymptote in performance.

Keywords: Absolute Identification; Modality; Set Size

In a typical AI task, stimuli vary on just one dimension, for example, line length or tone intensity. These stimuli (called unidimensional stimuli) are first presented to the participant labelled with a unique marker. The smallest is labelled # 1, the next # 2, and so on. In the test phase, the participant is presented with one item at a time, and asked to label it with its unique referent. An incorrect response is followed by the display of the correct stimulus label. Using this seemingly simple task, researchers have examined the capacity of human information processing by calculating the amount of information transferred from stimuli to responses, where an increase in information transfer is analogous to an increase in memory for stimulus items. Information transfer is measured in units of *bits,* taken from *information theory* (Attneave, 1959).

AI tasks provide a test of recognition for a set of unidimensional stimuli. Given the apparent simplicity of the task and the generally unlimited capacity of long term memory for stimuli such as faces, one would expect that a participant in an AI task would be able to reach perfect performance (in the least, if given enough practice). Unidimensional stimuli however, are an exception to the usually unlimited capacity of long term memory. Performance in tasks using unidimensional stimuli is generally very poor, and fifty years of research have consistently shown that even extensive practice does not significantly increase this limit in memory (Garner, 1953; Weber, Green & Luce, 1977; Shiffrin & Nosofsky, 1994).

This limitation in performance has become a truism of AI research: people cannot learn to improve their performance. Even significant practice has little effect on performance in AI tasks. For example, Pollack (1952) used absolute identification of tones varying in pitch, and found that after several days of practice, information transfer remained at only 2.3 bits, equivalent to perfect identification of only approximately 5 tones. In a similar experiment, Hartman (1954) found that during an eight week testing period, performance increased very slowly and never approached perfect levels. Garner (1953) concentrated on absolute identification of tones varying in *loudness* rather than pitch, and found even when using 12,000 trials, information transmission was still low (1.62 bits). Similar to Garner and Weber, Green and Luce (1977), also used 12,000 trials, and compared the performance on the initial and last 2,000 trials to calculate change in accuracy. Even with monetary incentives and significant practice, performance was only shown to improve 8.5%. Weber et al. only used six stimuli, which is fewer than the generally accepted short-term memory limit of 7±2. Even so, practice failed to significantly improve performance, and no ceiling effects were found.

More recently, Rouder, Morey, Cowan and Pfaltz (2004) demonstrated that learning is possible in a unidimensional AI task. Rouder et al. used an experiment with 30 lines varying in length, and found that accuracy improved significantly with only moderate practice. In stark contrast to Garner (1952) and Weber et al.'s (1977) large number of trials, Rouder et al.'s (2004) participants had, at most, only 15 hours of practice over a period of 10 days, or a maximum of only 7,200 trials. Two participants (including two authors of the research) performed three absolute identification experiments of different set sizes (13, 20 and 30 lines of increasing length), with an additional naïve participant in the third experiment only. Rouder et al. showed that not only were participants able to improve with practice, but performance showed no indication of reaching an asymptote. One participant even reached near-perfect performance within only five sessions in the 13 stimuli set, and within four sessions in the 20 stimuli set. Participants in the 30 line stimulus set also showed significant improvement over time: the average probability of correct responses increased by .28 over six sessions.

These results suggest that, contrary to previous literature, practice in Rouder et al.’s (2004) unidimensional AI task resulted in improved performance. It remains unclear which of the atypical features of Rouder et al.’s design were responsible for these results. AI tasks typically employ tones varying in intensity (e.g. Garner, 1953) or pitch (e.g. Hartman, 1954; Pollack, 1952) as stimuli, however Rouder et al. used lines of varying length. While some have used lines as stimuli (e.g. Lacouture, 1997; Baird, Romer and Stein, 1970), the use of tones is more standard. Rouder et al. also gave participants two response attempts for each trial. Traditional AI tasks give feedback after each trial, where participants are only given one response attempt. If this response is incorrect, the correct answer is displayed. In Rouder et al.’s experiment, however, if the first response was incorrect, the participant was given another response opportunity. If the second response was incorrect, the correct answer was displayed.

In order to begin the investigation into Rouder et al.’s atypical findings, we first replicated their third experiment to determine whether results are reproducible with naïve participants. Further to this, we investigated whether the feedback technique or stimulus modality influences any learning effect found.

Experiment 1: Lines Varying In Length

Participants

Twelve participants took part in Experiment 1, with six participants in each condition. Participants were reimbursed for their time and effort at $15 per session.

Stimuli

In both conditions, we used 30 lines of increasing length shown on a 21-inch CRT monitor, using 1152x864 resolution. See Table 1 for line lengths in pixels. Lines were displayed with 22x22 pixel jitter, to avoid the use of any visual cues.

Procedure

Each participant was first shown the stimuli in ascending order, labelled with a number from 1 through to 30. The text, “This is line number *n*”, was also displayed, and the participant was required to select the correct response to continue. Responses were made using a mouse and 30 buttons on the left hand side of the screen. The buttons were arranged in 3 columns of 10 buttons, with labels from 1 to 30 filled by row.

Participants completed ten sessions over ten, mostly consecutive, days. In the first three sessions, participants completed six blocks of 90 trials. On the following seven sessions, participants completed seven blocks of 90 trials, resulting in 201 presentations per stimulus.

The two conditions in Experiment 1 differed in their feedback technique. Experiment 1a used a feedback technique similar to Rouder et al. (2004), where participants were given two opportunities to respond. If the first response was incorrect, the participant was given another response opportunity. If the second response was again incorrect, the correct response was displayed. Experiment 1b used a more traditional feedback technique, where participants were only allowed one response opportunity. If the first response was incorrect, the correct response was displayed. In all cases where the response was correct, a pleasant 500ms tone was played.

Table 1: Line lengths used In Experiment 1

Length of lines in Experiment 1 in Pixels									
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

Analysis

Performance was measured in two ways: using accuracy and information transfer measures. In order to avoid over-estimation of performance, information transfer was analysed in ‘runs’, or pairs of sessions. For ease of comparison, we also report accuracy in the same ‘runs’.

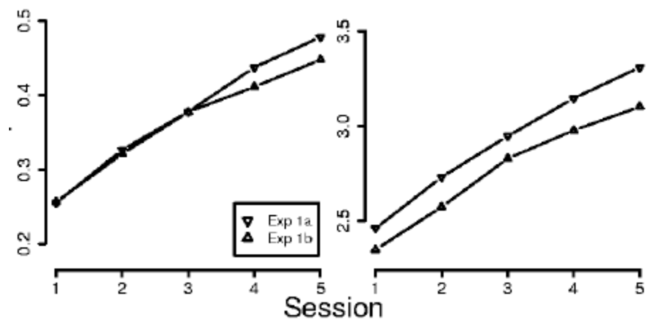
Results were analysed using the lme4 package (Bates, 2005) available for the statistical program, R. Maximum likelihood linear mixed effect binomial-probit regressions were estimated for accuracy data. Linear mixed effect models with random subject intercepts were also used for analysis of information data, assuming constant variance Gaussian error and restricted maximum likelihood estimation. Bayesian Markov Chain Monte Carlo Methods (see Baayen, Davidson & Bates, in press) were used to perform inference.

Results

Participants in Experiment 1a (two response condition), improved their accuracy by 22% and information transfer by

.84 bits over the length of the experiment. Participants in Experiment 1b increased their performance in a similar manner, where accuracy improved 19% and information transfer increased .76 bits (See Figure 1).

Polynomial contrasts were conducted on runs, accounting for differences in the number of trials in each. Reliable linear and quadratic effects were revealed for information transfer (p<.001 and p=.007) and probit accuracy (both p<.001) suggesting improvement across sessions. No reliable difference was found between average performance on the two response and one response conditions for either probit-accuracy (p=.86) or information transfer (p=.51). A significant interaction with the linear contrast for probit-accuracy however, showed that participants in the one response condition improved at a slower rate compared to participants in the two response condition (by 14%, p= 002). The same effect was not evident for information transfer (p=.30), but this could be a result of the power of information analysis.



A) B)

Figure 1. Population average estimates from linear mixed models for Experiment 1. A) Improvement in accuracy across five runs (where each run is equivalent to two sessions). B) Improvement in information transfer across the 5 runs.

The rate of improvement of information transfer was 9% slower in the one response condition, compared to the two response condition. This is equivalent to an increase across runs of .156 and .172 bits per 1000 trials respectively. When accounting for the increase in number of responses made to stimuli in the two response condition however, almost identical learning rates were found (an increase of .151 bits per 1000 for the two response condition, 3% smaller compared to the one response condition), suggesting that the increase is largely the result of an increase in the responses made for the two response condition.

Discussion

Experiments 1a and 1b confirmed the findings of Rouder et al. (2004), that learning is possible in a unidimensional AI task. Furthermore, we have confirmed that Rouder et al.'s results were not simply the product of exceptional participants. Experiments 1a and 1b demonstrate that the two response feedback system was not responsible for the learning effect, as when results are analysed based on responses, as opposed to trials, there is little difference in the rate of learning between the two.

Experiment 2: Removing External Cues

While the results of Experiment 1 suggest that learning is possible in AI if given sufficient opportunity to practice, it is possible that participants were using external cues to base their judgments of line length. For example, participants may judge line lengths relative to the edges of the computer monitor or some other physical cue. Such judgements were presumably less available in traditional experiments using auditory stimuli, which may explain the discrepancy between our experiment (and Rouder et al.'s 2004) and the traditional absence of learning observed in AI tasks. In Experiment 2, we remove these external cues by conducting the experiment in a dark room, as well as covering the edges of the monitor, masking the stimuli, and removing the response buttons when the stimulus is in view. It is believed that hiding the edges of the monitor and the response buttons would remove the potential of using these as relative size cues. In addition, the stimuli masks meant that no visual after image could be left on the screen that would allow participants to make comparative judgments.

Participants

Six participants took part in this experiment, and were reimbursed in a similar fashion to the participants in Experiment 1.

Stimuli

30 pairs of dots varying in separation were used in Experiment 2. Each set consisted of two dots, spaced apart at increasing intervals. Dots were used instead of lines with the intention of removing any effect of luminance. The separation of these dots was equivalent to the length of the lines used in Experiment 1 (see Table 1). Dots were white on a black background. The room was dark except for the illumination due to the computer monitor.

Procedure

Participants engaged in an AI task similar to that of Experiment 1, with the exception that participants in this experiment took part in the task in a dark room, where the only light provided was that which was given by the computer monitor In order to reduce the light given by the monitor, the background of the experiment was black, and the stimuli and other associated items were white. A cue (+) was shown before the presentation of a stimulus. A mouse click on the cue began the next trial. Stimuli were displayed on the black background for 1000ms, after which the response buttons appeared on the left hand side of the screen, and a mask of about 75 randomly scattered white dots covered the stimulus. Participants took part in six blocks per session, resulting in the presentation of each stimulus 180 times.

Results

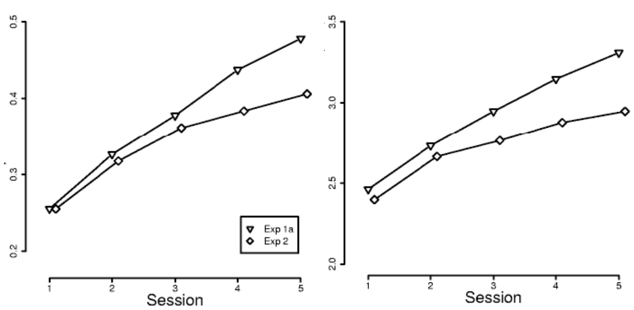
Participants in Experiment 2 increased their accuracy by 15% and information transmission by .55 bits across the 5 runs (See Figure 2).

There was no reliable difference in either accuracy (p=.66), or information transfer (p=.39), between Experiment 1a (30 lines with two response opportunities) and 2. There was a reliable interaction however, between Experiments 1a and 2 and the linear run contrast (p<.001), due to a decrease in the rate of learning. Participants in Experiment 2 learnt at a slower rate compared to participants in Experiment 1a (slower by 38% and 32% for bits and accuracy respectively).

Linear regression analyses conducted on information and probit accuracy separately for each individual's data, produced slopes that were reliably greater than zero, showing that all individuals improved their performance across the sessions. The rate of increase in information transmission for each individual ranged from .08 to .17 bits per 1000 trials, suggesting a slower learning effect compared to Experiment 1a. As with Experiments 1a and 1b, there was still no evidence of participants reaching an asymptote in their performance levels during practice.

Discussion

While the removal of cues appeared to slightly slow the learning effect, the effect observed was still much larger than found in previous research (e.g. Weber, Green & Luce, 1977). The rate of learning was slower for Experiment 2, but a substantial learning effect was still evident, suggesting that visual cues alone cannot account for the learning effect.



A) B)

Figure 2. Population average estimates from linear mixed models for Experiment 2 and Experiment 1a. A) Improvement in accuracy across five runs B) Improvement in information transfer across the 5 runs.

Summary of Results

The level of performance in our absolute identification tasks varied slightly, however all experiments showed evidence of significant learning. This is contrary to a vast amount of previous research that has shown little improvement in performance, even when given significant practice (e.g. Garner, 1953; Weber, Green & Luce, 1977). Weber, Green and Luce (1977) for example, used 12,000 trials and showed very little improvement in performance. In contrast, we used on average only 5820 trials and found evidence for significant improvement in performance for both lines varying in length and dots varying in separation.

A comparison with data from Dodds, Donkin, Brown and Heathcote (2008) also allow some insight into the role of set size on learning in AI. Dodds et al. (2008) describe an experiment using two conditions, where participants practiced with either 16 tones varying in intensity or 16 lines varying in length. They observed improvement with practice in both cases, but, unlike the current research, also demonstrated evidence of an asymptote well before the end of practice. In conjunction with the current research that uses larger set sizes, these results suggest that a large set size is important for supporting practice effects in AI, as while those in the 16 stimuli condition demonstrated improvement, they did not continue to improve across practice, and instead showed performance similar to that of which has been previously reported (e.g. Pollack,1952) This may be another reason why traditional experiments have not found learning effects.

Dodds et al. (2008) also found faster learning rates and greater asymptotic performance levels for participants who practiced with 16 lines of different length than those who practiced with 16 tones of different loudness. This suggests that stimulus modality or perhaps pairwise discriminability (the ability to distinguish between adjacent stimuli in the set), affects learning. A comparison between the performance of those in the 16 lines condition from Dodds et al. and those in the 30 line condition in the current experiment however, show that those in the 16 lines condition improved at a *faster* rate compared to those in the 30 lines condition. This suggests that while set size may affect the limitation in performance, it cannot be the *only* factor influencing the learning rate.

While all efforts were made to encourage consistency between experiments, during the course of the research, it was clear that motivation severely fluctuated between participants. It is therefore possible that motivation may constitute the high variability between subjects, and be responsible for the lower rate of learning found with smaller set sizes. Despite this seemingly plausible explanation however, the popular anecdote from Shiffrin and Nosofsky (1994) shows us that even the most highly dedicated participant may still have trouble in increasing their performance in a unidimensional AI task. Nosofsky, himself a dedicated AI researcher, locked himself in a soundproof booth for days, and yet still failed to achieve perfect, or even considerably improved, performance when practicing with tones.

Discussion

Our experiments suggest that thereare differences in the ability to learn unidimensional stimuli that could be due to modality (lines vs. tones) and individual motivation. In addition, when considering results from Dodds, Donkin, Brown and Heathcote (2008), a large set size also appears essential for participants to demonstrate learning of these unidimensional stimuli. Here, we propose an integrative hypothesis that draws these effects together, and reconciles our findings with previous research.

Figure 3 demonstrates the relationship between initial performance and overall improvement. Higher initial performance is associated with greater improvement in information transfer ($r(16)= 0.57$, $p=0.013$) across all experiments. This pattern also appears in Rouder et al.'s (2004) Experiment 3, where participants demonstrated significant learning with 30 lines of varying length when given moderate practice. Rouder et al. noted that learning varied as a function of initial accuracy (p. 939). They suggested that this was evident in Experiment 3 where all three participant's started with moderate accuracy (mean=.42) and increased to .70.

This correlation is even more surprising given that the naïve expectation would be for the opposite (a *negative* correlation), since those participants who begin with greater accuracy presumably have less headroom for improvement. This striking relationship thus forms a simple working hypothesis – that learning in AI will occur whenever a participant with a high initial performance level practices with a sufficiently large stimulus set. This hypothesis naturally explains the difference between stimulus modalities, such as Nosofsky's famous null finding. Learning cannot occur for tones of varying loudness because the perceptual variability observed for this dimension is such that a sufficiently large number of stimuli cannot be generated that are perfectly pairwise discriminable (without having tones that are both uncomfortably loud and impossibly quiet). Similarly, our hypothesis explains why early experiments using just a few stimuli (such as Weber et al.'s, 1977) did not observe learning – large set sizes are required. Explicit experimental tests of this theory are possible, and underway in our laboratory. For example, one could conduct an AI experiment similar to those above, and manipulate the spacing between stimuli. Using lines varying in length, one could manipulate the distance between stimuli between subjects, where one group practices with lines that are spaced more closely together than used in the current research, and another with lines that are spaced further apart. Our hypothesis would predict that more closely spaced stimuli would lead to poor pairwise discriminability and hence poor initial performance. This would lead to a failure to demonstrate learning.

The mechanism underlying our hypothesis however, is uncertain. Perhaps early positive reinforcement of stimuli and response allows greater consolidation of the information they acquire. Therefore while motivation cannot be the sole explanation of the differences in learning between experiments, perhaps greater motivation could lead to higher initial accuracy, simply by being more engaged and determined in the task early in the task, and hence allow greater improvement overall through early consolidation of information.

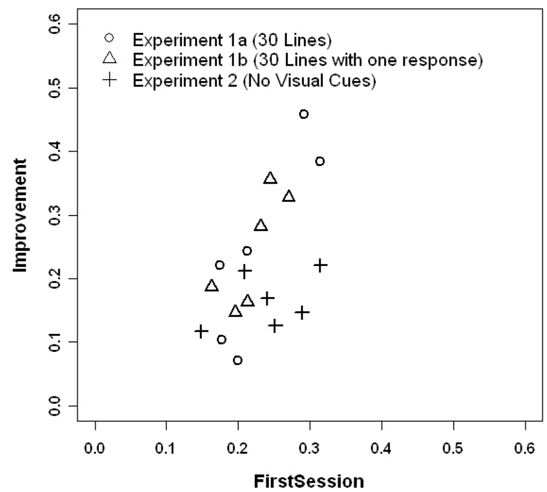


Figure 3. The relationship between initial performance and improvement in performance. Data shown is from each individual from all experiments. First Session data is from the first 540 trials. Improvement is the difference between the initial performance and final information transfer value (calculated based on last full 540 trials completed).

General Discussion

It appears that learning is possible in unidimensional AI tasks, and it is not the result of exceptional participants, or an abnormal feedback technique. These results are in stark contrast to a large amount of previous research that has suggested we are not able to improve our performance.

We suggest that while learning is possible in unidimensional absolute identification tasks, it is dependent on several factors. Participants must be given adequate opportunity to practice (taking into account the number of presentations of each stimuli, and not only the number of trials), and they must have moderate/high initial performance levels. The likelihood of initial accuracy being high may be enhanced if participants are sufficiently motivated. In addition, participants also need a large set size, and stimuli must be pairwise discriminable. Therefore it appears that learning is possible in unidimensional AI, but only under certain circumstances.

Acknowledgments

This research was partially supported by an ARC Discovery Project to Brown and Heathcote (DP0881244).

References

Attneave, F. (1959). *Applications of Information Theory to Psychology*. New York: Holt, Rinehart & Winston.

Baayen, R. H., Davidson, D. J. & Bates, D.M. (in press). Mixed-effects modeling with crossed random effects for subjects and items, *Journal of Memory and Language.*

Baird, J. C., Romer, D. & Stein, T. (1970). Test of a cognitive theory of psychophysics: size discrimination. *Perceptual Motor Skills*, *30*(2), 495-501.

Bates. D. M. (2005). Fitting linear mixed models in R. *R News*, *5*, 27-30.
Dodds, P., Donkin, C.M., Brown, S.D. & Heathcote, A. (2008). Practice effects in absolute identification: Breaking Miller's Limit, *Australian Journal of Psychology. 60(supp),* 68
Garner, R. W. (1953). An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology, 46*(5), 373-380.
Hartman, E. B. (1954). The influence of practice and pitch distance between tones on the absolute identification of pitch. *The American Journal of Psychology, 67*(1), 1-14.
Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits in our capacity for processing information. *The Psychological Review, 63*(2), 81-97.
Pollack, I. (1952). The information of elementary auditory displays. *The Journal of the Acoustical Society of America, 24*(6), 745-749.
Rouder, J. N., Morey, R. D., Cowan, N., & Pfaltz, M. (2004). Learning in a unidimensional absolute identification task. *Psychonomic Bulletin & Review, 11*(5).
Shiffrin, R. M., & Nosofsky, R. M. (1994). Seven plus or minus two: A commentary on capacity limitations. *Psychological Review, 101*(2), 357-361.
Weber, D. L., Green, D. M., & Luce, R. D. (1977). Effects of practice and distribution of auditory signals on absolute identification. *Perception and Psychophysics, 22*(3), 223-231.