Running head: NEIGHBORHOOD EFFECTS AND PRACTICE

Neighborhood Effects in Absolute Identification and the role of Practice Effects on Accuracy

Andrew D. Torr

Faculty Mentor: Scott Brown

University of California, Irvine

Abstract

The role of practice on neighborhood effects in absolute identification was investigated. Determining how overall proportion correct is affected by the size of the previous stimulus was a primary focus. Specifically, how the difference between the current stimulus presented and the previous stimulus affects overall accuracy was investigated. Participants were assigned a typical absolute identification task, judging line lengths over 800 trials. Each participant completed several sessions of trials over a period of a week. Practice increased overall proportion correct, but results are inconclusive.

The purpose of this experiment is to study a concept known as absolute identification. Miller (1956) was the first to discover this factor in perception. He noticed that subjects cannot judge differences in stimuli that vary only in one dimension with much accuracy. Absolute identification tasks are surprisingly hard for people to do, humans appear to be very bad at this type of task. In absolute identification tasks, the stimuli are simple and have a noticeable difference between neighboring stimuli, for example ten lines of gradually increasing lengths. Absolute identification involves identifying stimuli for any sense that vary along a single dimension. Ten sounds gradually changing in amplitude or pitch are other examples.

There are a few common things that occur in absolute identification tasks. First, stimuli used in absolute identification tasks are typically perfectly pairwise discriminable. This means that when only two stimuli are shown, they are never confused. This shows that it is not a perception problem. Stimuli can be judged relatively accurately with few stimuli, but the accuracy level then drops severely once there are more than 7-8 stimuli (Garner, 1953).

Surprisingly, this impairment in performance is evident no matter how big the difference between stimuli. If the stimuli vary greatly in size or have very little variance at all, for some reason, people still perform just as badly. A study was performed that examined whether the range of the stimuli set had an effect on performance (Pollack, 1952). It was discovered that the range of the stimuli set had little effect.

Accuracy seems to be higher with the end stimuli (Luce et al, 1982). For example, in the task assigned in this study, the longest and shortest lines would be the end stimuli. This is referred to as a bow effect. One possible explanation is proposed by

Berliner and Dulach (1973). They suggest that the end stimuli are used as “anchors,” or reference points to compare other stimuli to, making them easier to identify. It is also at least partly possible that since there are not alternate choices in both directions on end stimuli (for example, in the present experiment, there is no line 0 or line 11) these trials are harder to respond incorrectly because the number of alternate choices near to the correct response is fewer.

There are other phenomena that occur in absolute identification tasks, referred to as sequential effects. Every absolute identification experiment documented shows sequential effects (Luce et al., 1982; Holland & Lockhead, 1968); they play a central role in models of performance on AI (absolute identification tasks). Sequential effects occur when performance on a particular stimuli or trial is affected by those that came previously. Examples of sequential effects would be assimilation and contrast (Ward & Lockhead, 1970). Assimilation occurs when response is biased towards a stimulus that came previously. For example, the previous stimulus was a larger line, so, responses are biased towards this end of the stimulus set. Contrast occurs when response is biased away from a previous stimulus.

Performance has been shown to change with practice (Rouder et al., 2004). Overall accuracy increases, and response time decreases. The greatest effect is typically seen on stimuli in the middle of the range. Performance rises on end stimuli after practice, but not to the same degree.

This experiment will investigate an observed phenomenon in which subjects seem to be affected by the size of the difference between the current and previous stimulus. This is known as a neighborhood effect. Rouder et al. (2004) showed that when the current and previous stimuli are close to each other in terms of location within the range of stimuli, subjects' responses are more accurate. Accuracy gradually decreases as the difference between the stimuli increases. Accuracy (or proportion correct, as it is referred to in the current experiment's data) then rises again near the edge stimuli (see Figure 1). After much practice, however, there appears to be a reduction in this effect. The overall variability of accuracy in each category (distances between current and previous stimuli) diminished, indicated by a flattening out of the line after practice. This study will specifically investigate this effect, and will investigate whether there is a reduction in the effect after practice.

Some procedures were used in the Rouder study that were not repeated in the current experiment. First, only three subjects were used in Rouder et al. (2004). Also, after each trial, if incorrect, subjects were forced to retry the trial (the data from retries was not collected or analyzed, it was simply used to aid learning). This study looked to see if accuracy could be increased with lots of practice over a long time period, so subjects repeated the task several times over weeks and months. Each subject completed four sessions within the period of a week, and more subjects were tested. Subjects were not given a second chance on each trial.

The goal of this experiment is to investigate the phenomenon of neighborhood effects, and specifically investigate whether this effect can be reduced with practice. How is accuracy affected by the size of the previous stimulus? Will subjects rely less on the mechanism behind this effect after several sessions of practice? This experiment will investigate how overall proportion correct changes according to which stimuli came earlier, and how this changes with practice.

Methods

Participants

Participants were recruited from the UCI undergraduate subject pool. There were fourteen participants, and all were undergraduates at UCI. Each participant completed four one-hour sessions. They were paid $10 per session completed.

Apparatus

Participants were tested on desktop computers. Every computer had a 19" monitor and a standard keyboard and mouse. Stimuli varied length from 104 to 371 pixels (see Figure 2). The difference between neighboring stimuli was 15%. The thickness for all lines was 10 pixels. The lines were labeled based on their relative size, the smallest being "1" and the longest being "10." The position on the screen where the line was presented was randomized, in order to prevent participants from using the edge of the screen as a reference.

Design and procedure

Participants were instructed to guess the label for lines, presented one at a time. After their attempt, they received feedback. If a correct answer was given, a high pitched tone indicated a correct answer. If an incorrect response was given, a low pitched tone indicated this. Each participant completed ten blocks of trials per session, each block containing 80 trials, totaling in 800 trials per session per participant. Participants completed four sessions, within the period of a week. In order to motivate participants, they were told that their pay was based on their performance, when actually, the compensation was always $10 per session, regardless of performance. They were told if they performed well (in the top 33% compared to previous participants), they would receive the full $10, and would receive only $8 if not. All participants were informed of the deception at the completion of their last session, and paid $10 per session, regardless of performance.

Results

The results found in a preliminary analysis of the data indicate common effects found in absolute identification. As indicated by the graphs of overall proportion correct and response time (Figure 2), bowing effects were found. Subjects' proportion correct was highest and response time was quickest on the end stimuli (1-2 and 8-10 in Figure 3). Overall performance increased as subjects progressed from session one to session four. There was a general stabilizing of proportion correct across lines as subjects progressed from session one to session four. The most dramatic increase is between sessions is between sessions one and two, specifically on proportion correct for lines. The graph for response time does not indicate this same dramatic change from session one to two as in the proportion correct, the change is more or less constant.

Subjects showed an advantage in performance when there were 0 intervening stimuli (when the previous trial presented the same line), as would be expected (see Figure 4). This advantage diminishes as the number of trials since the last presentation of that line increases. An interesting phenomenon is found on trials when there are two intervening stimuli (three trials back). Proportion correct gradually decreases from 0 to 2 intervening stimuli, and then rises slightly after that. This effect is constant across all sessions.

Neighborhood effects were also found, indicated in Figure 5. Proportion correct was highest in all sessions on the trials in which the current and previous stimuli are close or identical, or are very far apart. The graph seems to indicate that there is a reduction in the neighborhood effects, illustrated by the flattening out of the line as subjects progress from session one to session four, especially between session one and the others. However, is this reduction in the variability simply what would be expected as proportion correct reaches the maximum? Can it be accounted for by ceiling effects?

The variability in the performance for each line must decrease as they approach the ceiling because the degree of variability is relative to the closeness to the ceiling, 100% correct. So, was the stabilizing of overall proportion correct accounted for by neighborhood effects or ceiling effects? What would happen if session one's overall performance were increased to that of session four, without changing the performance in any other way? To analyze this, linear scaling was applied. We applied a formula to the proportion correct for each line that increased the overall proportion correct of a session to that of a session of our choosing. This means that we adjusted the overall proportion correct of the earlier session to match that of the later session. The graphs in Figure 6 incorporated the method described. As indicated by the graphs, when overall performance on any session is increased to that of the following session, the lines appear virtually identical. The most dramatic change would be expected between sessions one and four, and again, the lines appear virtually identical. This indicates that the stabilizing of the data can almost entirely accounted for by ceiling effects.

In order to understand why supporting evidence for the hypothesis was not found, an analysis and comparison of the data from the study performed by Rouder et al. (2004) is necessary. The same method of linear scaling was applied to Rouder's data (see Figure

6). The data shows a reduction in neighborhood effects, indicated by a general stabilizing of the After Practice data across categories, as compared to the Before Practice data.

Discussion

Analysis of subjects' overall performance shows typical effects found in absolute identification. Bow effects were found in an analysis of proportion correct and response time on each line. Subjects performed more accurately and faster on trials in which there were zero intervening stimuli. One or more intervening stimuli eliminated this advantage.

Typical neighborhood effects were found in the analysis of subjects' performance, when categorized by the difference between the current trial and previous trial. Additionally, a reduction in the variability of the data between such categories was found, indicating a reduction in neighborhood effects. However, as indicated by the linear scaling analysis, the reduction was no larger than would be expected by increased proportion correct. While comparing performance on early sessions to that of later sessions in an absolute sense shows the appearance of a reduction in neighborhood effects, this apparent reduction is actually due to ceiling effects. When data from a previous study (Rouder et al., 2004) was subjected to the same analysis, the reduction in variability cannot be accounted for by ceiling effects alone, there is a reduction in neighborhood effects.

Why was there no reduction in neighborhood effects in the presented study? There are a few possibilities. There are some methods that were used in Rouder that were not used in this experiment. First, the subjects used in the previous study were tested over several weeks while the presented experiment tested subjects over a period of no more than one week. It is possible that more practice is needed in order to show a noticeable reduction in neighborhood effects.

Another difference in Rouder's study that was not used in this experiment involves forcing subjects to try again on trials where they responded incorrectly. Making subjects pay closer attention to their errors might help them not make the same mistakes. If this were true, it would likely have the strongest effect on the lines in which the most errors occurred, resulting in a reduction in the variability of the proportion correct for each line, and therefore, a reduction in neighborhood effects.

An additional difference between the current experiment and Rouder's study is the number of stimuli. Rouder used 30 lines, while only 10 were used in this experiment. This results in a dramatic increase in the number of categories of data for the same effect. Since only 10 lines were used in this study, it may be harder to notice a difference in a particular set of categories. For example, notice that the graph of Rouder's data ranges from -8 to 8. This is only the center section of the entire spectrum of data. If the same focus of the current experiment's data were presented, it would only range from -3 or -2 to 2 or 3. Having more categories of data to analyze the same section of the data may make showing an effect possible.

One more possibility that could have affected the results is subject motivation. Two of the three subjects used were two of the authors of the study performed by Rouder. These two subjects would be highly motivated to learn to do the task well, and probably much more motivated than the average paid subject. All subjects in the current study were compensated monetarily. All three subjects in the previous study are claimed to be highly motivated. While there is no evidence that subjects were not motivated, no claims as to the degree of subject motivation can be made in the present study. The lack of personal involvement may strongly affect the degree of learning. Additionally, while they were informed that their compensation amount would be dependant on their performance, an additional $2 per session may not be enough motivation to seriously try to learn the task. They may have been content to do the task with minimal effort. A possible remedy to this would be to adjust the pay system. A large difference in compensation between the low performance group and the high performance group (perhaps $7 and $13) would be a suggestion to resolve this. Incorporating a point system may help as well, especially if an element of competition could be added. Rating subjects based on their performance compared to their last session would be another option.

Although no reduction in neighborhood effects was found in the current experiment, it is not entirely without benefit. Since a previous study showed the reduction in neighborhood effects while the current experiment did not, it is most likely one of the differences listed that has a large impact on this reduction. It would be recommended to attempt the experiment again, incorporating one or more of the differences listed above. Perhaps then the reduction in neighborhood effects found in Rouder et al. (2004) can be found, lending evidence to the hypothesis presented here. Fully understanding sequential effects is essential in order for researchers to create accurate models of subject performance in absolute identification tasks.

References

Berliner, J. E., & Durlach, N. I. (1973). Resolution in roving-level discrimination. *Journal of the Acoustical Society of America, 11*(5), 1270-1287.

Garner, W. R. (1953). An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology, 46*(5), 373-380.

Holland, M. K., & Lockhead, G. R. (1968). Sequential effects in absolute judgements of loudness. *Perception and Psychophysics, 3*(6), 409-414.

Luce, R. D., Nosofsky, R. M, Green, D. M., & Smith, A. F. (1982). The bow and sequential effects in absolute identification. *Perception and Psychophysics, 32*(5), 397-408.

Miller, G. A. (1994). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review, 101*(2), 343-352.

Pollack, A. (1952). The information of elementary auditory displays. *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), 938-944.

Ward, L. M., & Lockhead, G. R. (1970). Sequential effects and memory in category judgments. *Journal of experimental psychology, 84*(1), 27-34.

Figure Captions

Figure 1. This graph displays false data to give an example the theoretical ideal in performance and neighborhood effects, and how it is hypothesized that performance would change after practice.

Figure 2. Lines used in experiment. These lines are not the exact length of the lines used in the study, but are proportional in their relation to each other. Lines were labeled from 1 to 10, 1 being the shortest and 10 being the longest.

Figure 3. Average proportion correct and response time as a function of stimulus length are shown (ordinal values). The top graph and corresponding y-axis display proportion correct and the bottom graph and y-axis show response time (in seconds). The x-axis displays ordinal line lengths. Data for all four sessions are displayed.

Figure 4. Average proportion correct and response time as a function of number of intervening variables. The top graph and corresponding y-axis display proportion correct and the bottom graph and y-axis show response time (in seconds). 0 intervening variables indicates that the previous trial displayed the same line, 1 intervening variable indicates that there was 1 trial in between current trial and the most recent presentation of the same line, 2 intervening variables indicates that there were 2 trials in between the current trial and the most recent presentation of the same line, and so on.

Figure 5. Proportion correct as a function of the difference between current and previous stimulus. Negative numbers (on the x-axis) indicate that the lines presented in the previous trials were longer than the line of the current trial, and positive numbers indicate that the lines of the previous trial were shorter than that of the current trial. The assigned label (a number from 1 to 10) of the previous trial's line was subtracted from the label of the current trial's line.

Figure 6. Graphs incorporating linear scaling are shown. Each graph displays proportion correct as a function of the difference between the current and previous stimulus for a particular session. A previous session (to that of the session displayed) is displayed, and scaled to have the same overall proportion correct to that particular session. The formula used to apply linear scaling is displayed below.

Figure 7. Graph applying linear scaling to data from study performed by Rouder et al. (2004). Before practice data has been scaled to have same overall proportion correct as after practice data.

Figure 1

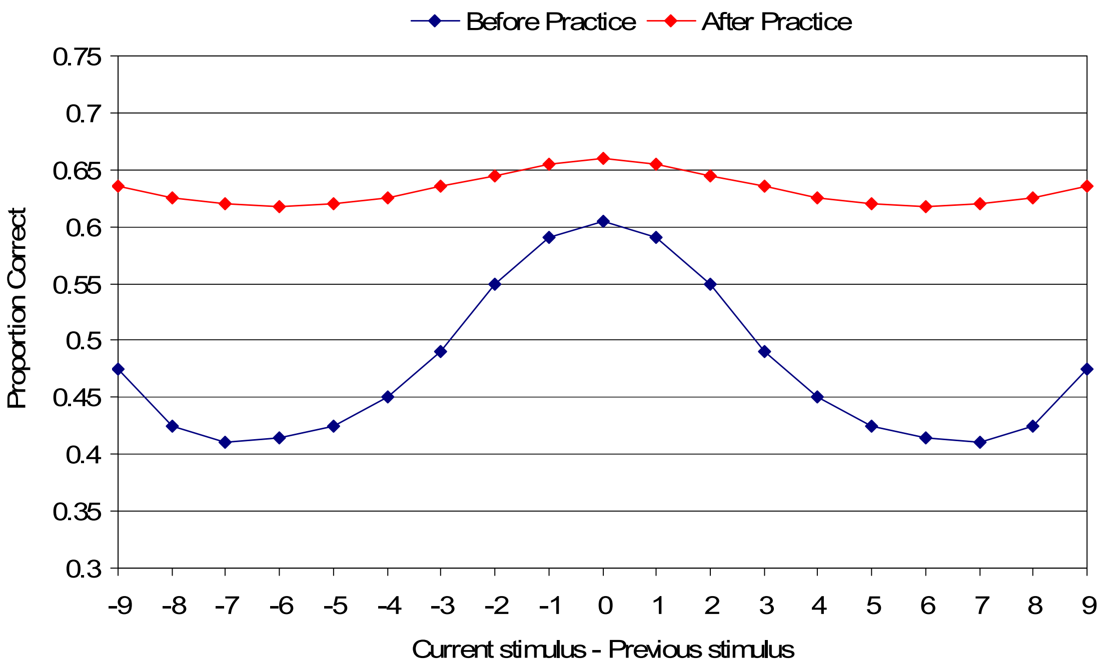

Before Practice
After Practice
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
Proportion Correct
-9
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
9
Current stimulus - Previous stimulus

Figure 2

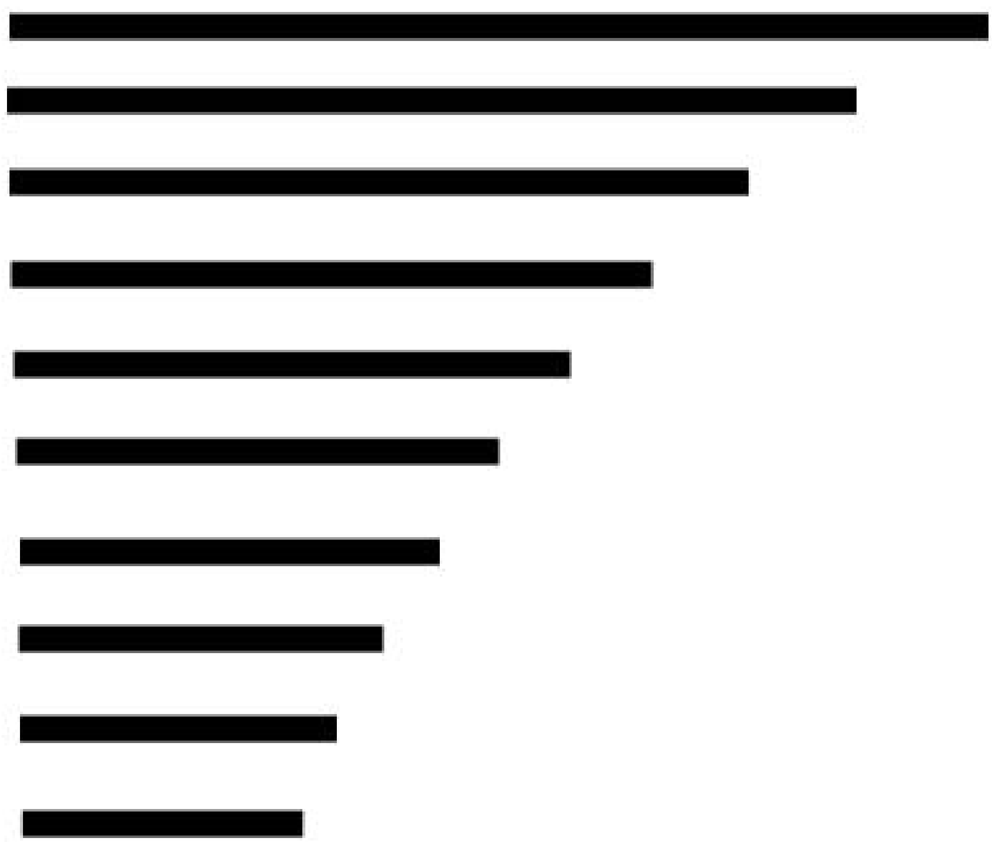


Figure 3

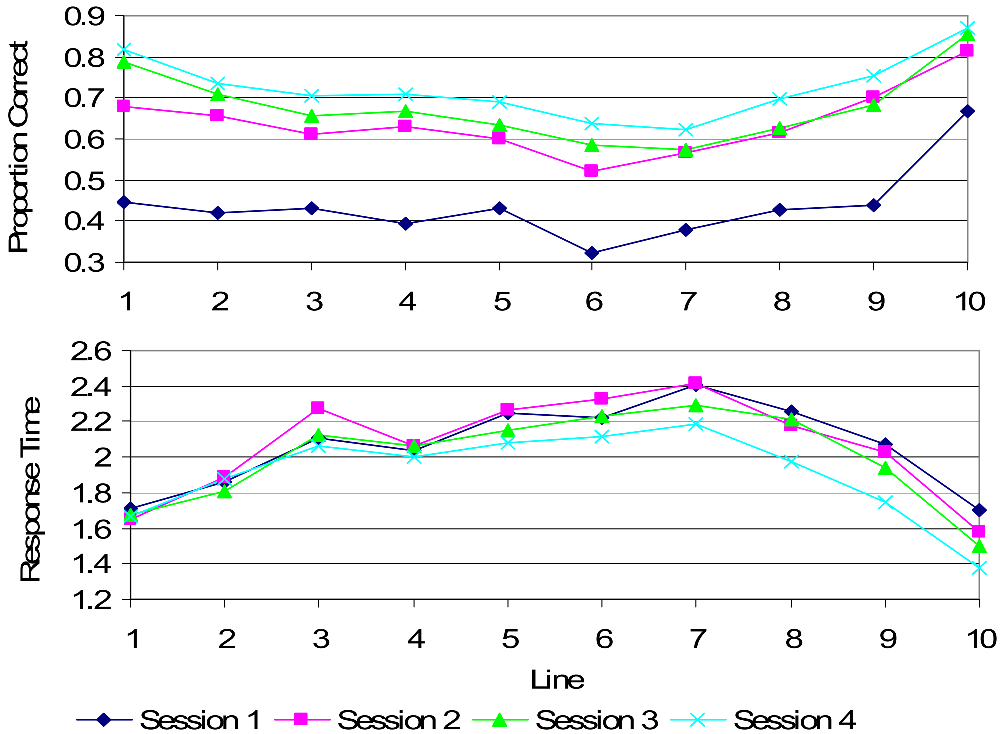

Proportion Correct
0.9
0.8
0.7
0.6
0.5
0.4
0.3
1 2 3 4 5 6 7 8 9 10
Response Time
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1 2 3 4 5 6 7 8 9 10
Line
Session 1
Session 2
Session 3
Session 4

Figure 4

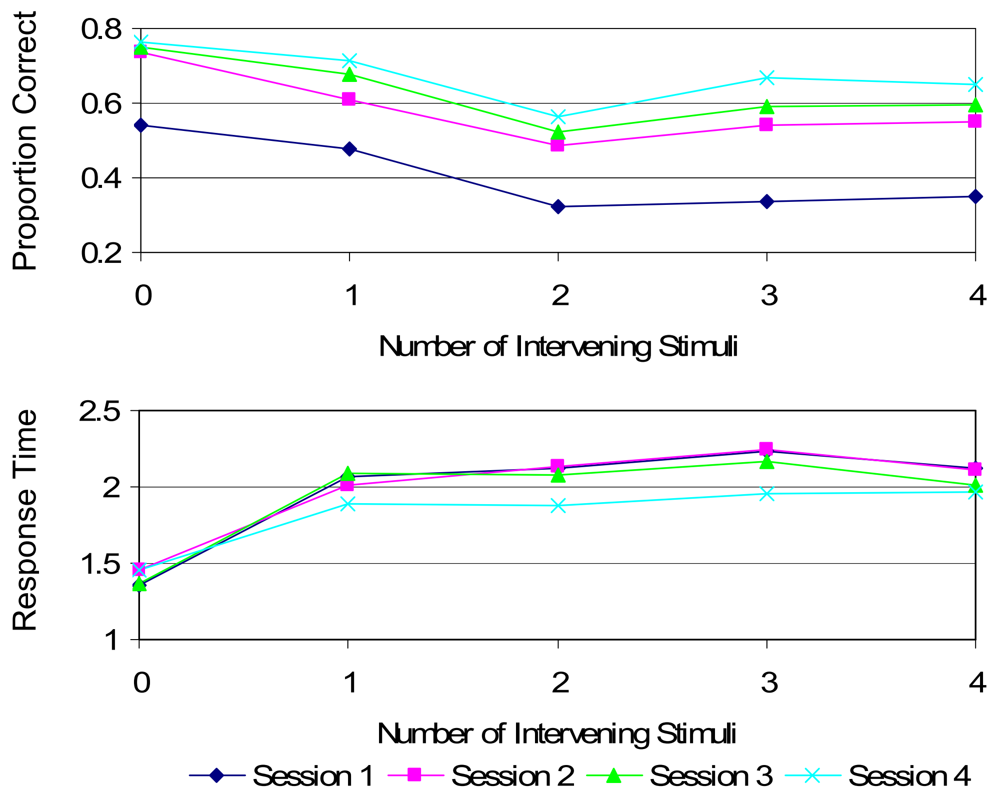

Proportion Correct
0.8
0.6
0.4
0.2
0
1
2
3
4
Number of Intervening Stimuli
Response Time
2.5
2
1.5
1
0
1
2
3
4
Number of Intervening Stimuli
Session 1
Session 2
Session 3
Session 4

Figure 5

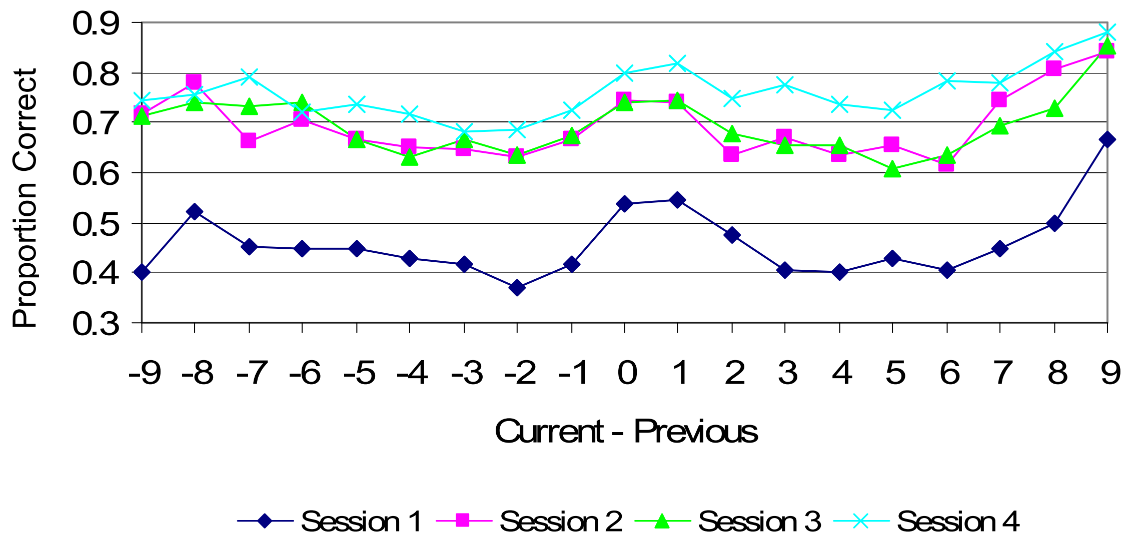
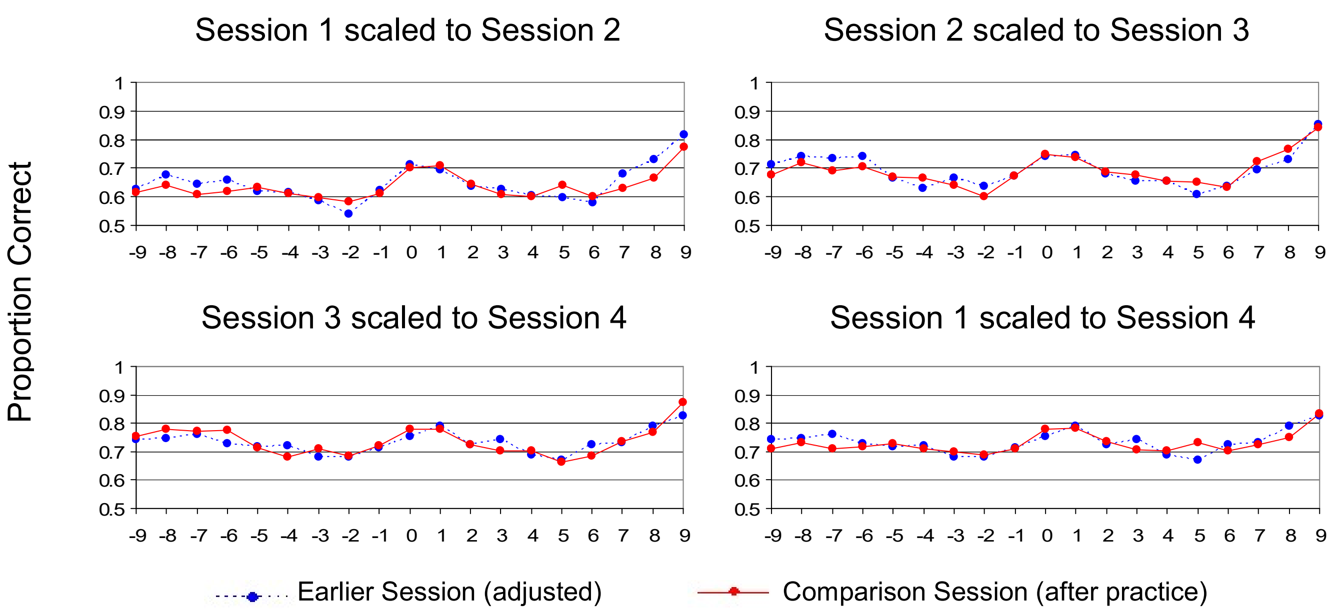
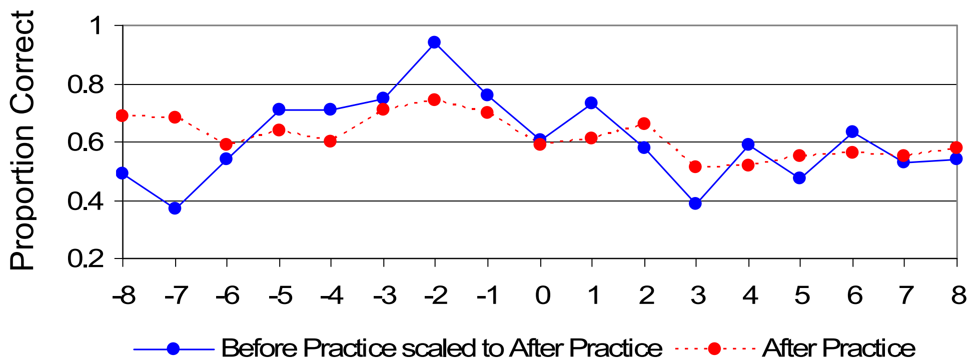

0.9
0.8
0.7
0.6
0.5
0.4
0.3
Proportion Correct
-9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9
Current - Previous
Session 1
Session 2
Session 3
Session 4

Figure 6



$$Y_{Adjusted} = 1 - \left(\frac{\text{Mean of Earlier Session - 1}}{\text{Mean of Compared Session - 1}} \left(1 - Y_{Compared} \right) \right)$$

Figure 7


Proportion Correct
1
0.8
0.6
0.4
0.2
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
Before Practice scaled to After Practice
After Practice