Effects of Inaccurate Feedback on Simultaneous Amplitude and Frequency Discrimination Tasks

Sophia Navarre
Rajeev Dutta
Mackenzie Wood
Francisco Reyes

Humans frequently engage in tasks with the intention of performing well. Many previous studies have explored feedback as a tool to improve this task performance, concluding that feedback improves task performance, but only when coupled with a task performance goal. However, there is no clear consensus on how feedback without goalsetting impacts task performance. This paper aims to investigate feedback’s impact on task performance from a new angle by determining how incorrect feedback affects task performance. Through this investigation, we explored the participant thought process that plays into how the participant uses feedback. This has several implications, for example, in regard to malingering, in which people purposefully provide incorrect feedback for personal gain. To conduct this study, we instructed participants to complete simultaneous tactile discrimination tasks on which they received correct or incorrect feedback. Our results showed a significant difference in task performance between correct and incorrect feedback, and this difference increased for a harder task. Additionally, our results demonstrated that several participants receiving incorrect feedback had longer response times, indicating that they realized they were receiving incorrect feedback. Overall, this study sheds light on how feedback can be used to manipulate task performance.

Introduction

Feedback is the delivery of information to the participant based directly on their performance. Learning is a key aspect of feedback because it is presumed the participant will use the feedback and apply it to the next task - thus having a direct impact on performance. The amount learned is also associated with the confidence the participant has in the feedback and their answers before feedback. Kulkavy et al [1] conducted a study focusing on feedback and the confidence participants held in their answers. It was found that feedback served as a good corrective measure whenever confidence is high, but not as much when confidence is low [1]. Kulhavy et al proposed that the reason feedback plays such an integral role is that it caused participants who responded confidently, but incorrectly, to thoroughly analyze their reasoning and to correct any misconceptions they had [1]. In addition, it was thought that confident correct answers increased the likelihood of answering similar questions correctly in the future.

Several studies in the late 1960s demonstrated that a subject’s motivation to do a task did not increase when they were provided with feedback on the task [2] Several studies in the late 1970s showed and confirmed that a subject’s performance on a task improved only if they had specific, hard goals and were provided with feedback. In 1981, a study showed that when subjects were each given an easier and a harder goal, they exerted more effort to attain the harder goal than the easier goal [3]. While these findings provided illuminating information about task motivation and performance, they did not determine how feedback affects task performance. Matsui and Okada noticed this and in 1983, they conducted two studies to test how feedback affects task performance. They did this by interrupting participants halfway through solving simple math
problems, allowing them to realize how many problems they had left so that the researchers could measure if this intermittent feedback influenced the second half of the problem-solving. Subjects were divided into a higher-scoring and lower-scoring group based on a 5-minute practice trial. In the experimental trial, they were told they had 10 minutes to solve 70 problems; this was described as a reasonable goal and 39.1% ended up achieving it. However, the subjects were stopped after 5 minutes (halfway point) and told to count the number of problems they had attempted. They then proceeded with the remaining 5 minutes. The mean number of attempts between the first and second part only increased for the lower-scoring subjects. Thus, this study found that feedback improved task performance only for subjects who felt they were behind halfway through the trial. Subjects in a control group were not interrupted at the halfway point. Regardless of if they were low or high scoring, their number of attempted questions was the same for both halves. Therefore, this study demonstrates that receiving feedback (i.e., knowledge of goal progress) is necessary for goals to affect task performance. The aforementioned math problem solving study as well as the studies it was based on have served as references on how to construct feedback experiments. Since they have been published, a majority of the papers that cite them are focused more on goal setting and less on feedback. This study differentiates itself because it focuses on feedback and does not include goal setting in the experimental design. Additionally, this study sets itself apart by exploring the impact of incorrect feedback.

To incorporate opposite feedback into tasks, the possible options for feedback can only be "correct" and "incorrect." In addition, the tasks should have a certain level of difficulty so that there is room for the participant to doubt their performance. For these reasons, this study had participants perform simultaneous tactile discrimination tasks. These tasks involved the use of a Brain Gauge, which delivers two slightly different stimuli to the ring and middle finger (digits 2 and 3). The instructions for the task are to press the button corresponding to the larger stimulus. With standard feedback, pressing the button corresponding to the larger stimulus results in a positive message. With opposite feedback, pressing the button corresponding to the larger stimulus results in a negative message. In this case, a correct answer could either be defined in two ways. First, a correct answer could be defined as choosing the larger stimulus. Based on the feedback message this would be considered incorrect, but based on the task instructions, this would be considered correct. Second, a correct answer could be defined as choosing the smaller stimulus. Based on the feedback message this would be considered correct, but based on the task instructions, this would be considered incorrect. Is failing to properly discriminate between two amplitudes actually failing if the feedback says it is correct? Is succeeding in properly discriminating actually succeeding if the feedback says it is incorrect? The answers to these questions certainly have real-world implications: for example, in the case of malingering, a purposeful attempt to feign symptoms to get insurance money.

Another study completed by Omrani et al. [4] investigated the effect of biased feedback on the perception of tactile stimuli and the ability to discriminate between different frequencies. The results of this study suggest that feedback played a major role in influencing the perception of tactile stimuli. As shown, many studies have been conducted to determine how receiving feedback affects performance in visual, tactile, and auditory tasks. This research on how feedback affects performance discrimination tasks could play a role in learning how to increase our brain’s ability to discriminate against frequency and amplitude. This study will focus on incorrect feedback that is hypothesized to be detrimental to learning. For this study, incorrect feedback was incorporated into frequency and amplitude discrimination tasks utilizing the Brain Gauge.

**Methods**

42 healthy individuals (aged 20-22) participated in the study. Participants completed amplitude and frequency discrimination tasks using Brain Gauge Pro (Cortical Metrics, Carrboro, NC). The program was run on participants’ personal computers, and Brain Gauge hardware was provided to each participant. The Brain Gauge hardware has two, five millimeter-diameter cylinders that are
capable of administering vibrotactile stimulations in the range of 25-50 Hz to the index and middle finger (through each cylinder). The Brain Gauge software provides instructions to the participants on the tasks they are to complete. This study tested two discrimination tasks: amplitude and frequency. For each, participants were provided two simultaneous vibrations (one from each cylinder on the Brain Gauge) of varying amplitude or frequency. Participants were then instructed to press the tip of the Brain Gauge hardware that delivered the more intense stimulus (i.e., the greater amplitude or the greater frequency). Each testing battery has a training phase, during which participants are provided feedback on whether or not they were able to successfully determine which tip delivered the more intense stimulus. If the participants are successful, the test becomes more difficult as it continues (i.e., the intensity of the simultaneous stimuli from each tip become increasingly similar to each other).

Participants were randomly assigned (single-blind) to either an accurate feedback condition (n = 23) or an inaccurate feedback condition (n = 19) that applied throughout the entire test battery. Under the accurate feedback condition, participants received either (1) a message indicating that they responded correctly, if they correctly identified the more intense stimulus or (2) a message indicating that they responded incorrectly, if they failed to identify the more intense stimulus. Under the inaccurate feedback condition, participants received either (1) a message indicating that they responded incorrectly, if they correctly identified the more intense stimulus or (2) a message indicating that they responded correctly, if they failed to identify the more intense stimulus. In other words, the accurate feedback condition provides veridical feedback on the participant’s performance, whereas the inaccurate feedback provides systematically misleading feedback on the participant’s performance. Data was collected on the success of participants in each group (e.g., speed and accuracy).

**Results**

Figure 1 shows the mean DL values for the amplitude discrimination task in which participants were given either correct, incorrect, or no feedback. This table shows the means calculated when outliers were removed as well as means calculated with the outliers still included. The DL values collected from participants who received correct feedback on the discrimination task did not contain any outliers and the mean value is 56.43 μm. For participants who did not receive feedback on the amplitude discrimination task, the mean DL with outliers is 57.65 μm and without outliers is 53.69 μm. A t-test between the correct feedback and no feedback data points shows that there is no statistical significance between their means, regardless of the inclusion or exclusion of outliers. The mean DL from the participants doing the amplitude discrimination task with incorrect feedback was 142 μm with outliers and 78.63 μm without outliers. A t-test between the incorrect feedback data and the correct feedback data yielded different results, depending on whether the data with outliers or without outliers was compared.

![Figure 1. Difference Limen and Response Time Value for Amplitude Discrimination](image)

The results of the t-tests are shown in Figure 2. A t-test between the incorrect feedback data and the correct feedback with outliers yielded a p-value of 0.017, indicating statistical significance. A t-test between the incorrect feedback data and the correct feedback without outliers yielded a p-
value of 0.19, which indicates that there is no statistical significance. The differences in the data for each type of feedback are depicted in Figure 1 (outliers included).

<table>
<thead>
<tr>
<th>Outliers Included?</th>
<th>Correct vs Incorrect Feedback</th>
<th>Correct vs No Feedback</th>
<th>Incorrect vs No Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p-value</td>
<td>statistically significant?</td>
<td>p-value</td>
</tr>
<tr>
<td>Yes</td>
<td>0.0166</td>
<td>yes</td>
<td>0.8965</td>
</tr>
<tr>
<td>No</td>
<td>0.1185</td>
<td>no</td>
<td>0.7260</td>
</tr>
</tbody>
</table>

Figure 2. Comparing Types of Feedback for Amplitude Discrimination

Figure 3 displays the mean DLs for a frequency discrimination task that different participants received different types of feedback for. The means were calculated with and without outliers. The mean DL for correct feedback were 4.15 Hz with outliers and 3.84 Hz without outliers. For incorrect feedback, the mean DL was the same both with and without outliers, at 9.69 Hz. For no feedback, the DL mean with outliers was 4.09 Hz and without outliers was 3.74 Hz.

<table>
<thead>
<tr>
<th>Correct Feedback</th>
<th>Incorrect Feedback</th>
<th>No Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference Limens (Hz)</td>
<td>Response Time (ms)</td>
<td>Difference Limens (Hz)</td>
</tr>
<tr>
<td>Outliers Included?</td>
<td>Mean</td>
<td>SD</td>
</tr>
</tbody>
</table>

Figure 3. Difference Limen and Response Time Values for Frequency Discrimination

Figure 4 shows the p-values that result when comparing data from participants receiving different types of feedback. The p-values of 0.92 and 0.85 (with and without outliers, respectively) indicate that there is no statistical significance between correct and no feedback data. Regardless of the inclusion or exclusion of outliers, there is statistical significance between the correct and incorrect feedback data, as well as between incorrect and no feedback. Without outliers, the p-value for correct vs incorrect feedback is 0.0028, whereas for incorrect vs no feedback, it is 1.024•10^-4. While both of these p-values are below the conventional alpha value of 0.05, the p-value for the comparison between the incorrect and no feedback data is one order of magnitude less. Similar to the DL means for amplitude discrimination, the incorrect feedback mean is higher than both the correct feedback and no feedback mean. This is depicted clearly in Figure 6. When comparing Figure 5 to Figure 6, one can observe that the “peak” mean corresponding to the incorrect feedback data is higher up in Figure 6 than in Figure 5.

<table>
<thead>
<tr>
<th>Outliers Included?</th>
<th>Correct vs Incorrect Feedback</th>
<th>Correct vs No Feedback</th>
<th>Incorrect vs No Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p-value</td>
<td>statistically significant?</td>
<td>p-value</td>
</tr>
<tr>
<td>Yes</td>
<td>0.0028</td>
<td>yes</td>
<td>0.9208</td>
</tr>
<tr>
<td>No</td>
<td>0.0018</td>
<td>yes</td>
<td>0.8479</td>
</tr>
</tbody>
</table>

Figure 4. Comparing Types of Feedback for Frequency Discrimination
Figure 5 shows a box and whisker plot for the DLs of amplitude discrimination tasks for participants receiving three types of feedback. The figure illustrates the lack of difference in mean between the correct feedback and no feedback data sets. The figure also illustrates the major difference in mean between correct and no feedback and incorrect feedback. The red plusses represent outliers, and Figure 5 shows that receiving incorrect feedback resulted in more outliers than receiving the other two types of feedback.

**Figure 5. Amplitude Discrimination Difference Limens with Different Types of Feedback**

Figure 1 shows the means response times of participants doing amplitude discrimination tasks. With outliers, incorrect feedback had the highest mean response time of 531.25 ms, no feedback had the second highest mean response time of 528.85 ms, and correct feedback had the lowest mean response time of 410.96 ms. Without outliers, no feedback had the highest mean response time of 441.65 ms, incorrect feedback had the second highest mean response time of 326.34 ms, and correct feedback had the lowest mean response time of 322.65 ms. The response time data is plotted against the corresponding DL data in three different figures, with each figure containing data from a specific type of feedback. Figure 7 shows the relationship between response time and amplitude discrimination DL for correct feedback, Figure 8 for incorrect feedback, and Figure 9 for no feedback. Figure 10 shows all of these together.
Figure 6. Frequency Discrimination Difference Limens with Different Types of Feedback
Figure 7. Amplitude Discrimination with Correct Feedback: Difference Limen vs Response Time. Large black circle corresponds to the mean Difference Limen and the mean Response Time.
Figure 3 displays the response times that participants had when completing the frequency discrimination tasks. With outliers, the mean response time for no feedback was 589.81 ms, for incorrect feedback was 464.77 ms, and for correct feedback was 372.37 ms. The way that the response times correspond to the DLs is shown in Figure 9-12. Figure 12 shows that the incorrect feedback data has more outliers in regard to the DL, but that the no feedback data has more outliers in regard to the response time.
**Figure 9.** Amplitude Discrimination with No Feedback: Difference Limen vs Response Time. Large black circle corresponds to the mean Difference Limen and the mean Response Time.
Figure 10. Amplitude Discrimination with Different Types of Feedback; Difference Limen vs Response Time
**Figure 11.** Frequency Discrimination with Correct Feedback: Difference Limen vs Response Time. Large black circle corresponds to the mean Difference Limen and the mean Response Time.
**Figure 12.** Frequency Discrimination with Incorrect Feedback: Difference Limen vs Response Time. Large black circle corresponds to the mean Difference Limen and the mean Response Time.
Figure 13. Frequency Discrimination with No Feedback; Difference Limen vs Response Time. Large black circle corresponds to the mean Difference Limen and the mean Response Time.
Discussion

As mentioned in the Results, Figure 1 shows that the order of slowest to fastest response time for amplitude discrimination changes when outliers are included or not. With outliers, the order (from slowest to fastest) is incorrect feedback, no feedback, correct feedback. Without outliers, the order is no feedback, incorrect feedback, correct feedback. It is important to note that the no feedback testing was done about 2 months before the other testing. During these 2 months, participants were completing other discrimination tasks with the Brain Gauge. This interim “practice” likely played a role in reducing the response times for all discrimination tasks, independently of the type of feedback. Since the mean DL of the no feedback and correct feedback tests were comparable, one could suggest that in the 2-month period the participants did not improve their discrimination abilities. However, they did get faster at making these decisions.

Figure 8 and Figure 12 illustrate the fact that participants receiving incorrect feedback took three different “approaches” to the discrimination tasks. The third quadrant contains data of a quick response time and small DL. These participants could have realized they were being given the wrong feedback so that they would just invert their answer. This would require skill in the discrimination tasks as well as confidence that you have “figured out” the test. This is plausible because the short response times indicate that these participants didn’t have any doubts with their answers. A less likely scenario is that these participants did not “figure out” the test, but rather
that the incorrect feedback did not affect them. This also indicates confidence and tracks with the short response times.

The second quadrant shows data points with a longer response time and a small DL. While ultimately these participants were able to successfully choose the larger stimulus, they took longer to contemplate their answer choice. For reference, the average response time is 500 msec on an amplitude discrimination battery and 500 msec on a frequency battery.

The first quadrant has no data points. It would correspond to a long response time and a large DL. To have a large DL, the participant must consistently choose the smaller stimulus, which would result in positive (incorrect) feedback. After consistently choosing the smaller stimulus, the discrimination task becomes easier, to a point where it is obvious which stimulus is larger.

The fact that there are no data points in the first quadrant indicates that of the participants who spent a longer than average time in selecting an answer, none sought to obtain a positive response from the computer over selecting what they thought to be correct.

The fourth quadrant has data points with a large DL and a short response time. This could mean that these participants were seeking to receive positive feedback from the computer. Whether this was intentional or not cannot be said conclusively. But since the short response time does indicate confidence, it seems that these participants were being intentional, as they were not taking time to struggle over figuring out which stimulus is greater. It is also possible that these participants were simply not good at discrimination tests, however this is unlikely since the DL values are much larger than those associated with the tests with correct or no feedback. As mentioned in the Results, the mean DL from the participants doing the amplitude discrimination task with incorrect feedback was 142 μm with outliers and 78.63 μm without outliers. This difference in means can be attributed to three participants whose DLs were above 400 μm, those in the fourth quadrant of Figure 8. The reason why it is important to consider the data with outliers is because of the strange nature of the incorrect feedback test. One could argue that a large DL actually shows success, because a large DL means that the participant was frequently choosing the lower stimulus, but then getting a message saying this was correct.

Similar to Table 1, Table 3 shows that removing outliers changes the frequency discrimination means. However, there were no outliers to remove for incorrect feedback frequency discrimination. This is quite the opposite of the incorrect feedback amplitude discrimination, just discussed. Nonetheless, Figure 12 still shows a fourth quadrant with 3 DLs higher than most of the other points. In this case, it is important to refer to Figure 5 and Figure 6, which illustrate the fact that even though within the DL data from the incorrect feedback frequency discrimination tests does not contain outliers, its DLs relative to the DLs for correct and no feedback are more pronounced than that of amplitude discrimination. Basically, the participants who had the least success in discriminating between frequencies were not outliers because the average participant did not have much success in this. Overall, Figure 5 and Figure 6 show that the trend of incorrect feedback yielding a larger DL is maintained throughout different discrimination tests. Their difference can be attributed to the different difficulty of the tests. Similar to the DL means for amplitude discrimination, the incorrect feedback mean is higher than both the correct feedback and no feedback mean. This is depicted clearly in Figure 6. When comparing Figure 5 to Figure 6, one can observe that the “peak” mean corresponding to the incorrect feedback data is higher up in Figure 6 than in Figure 5. One reason for this could be the level of difficulty of the discrimination tasks. Historically, the frequency discrimination task has been more difficult for participants. This is because participants have to learn what a frequency feels like, whereas the feeling of an amplitude is more intuitive.

As mentioned in the Results, Table 4 demonstrates that there is statistical significance between the correct and incorrect feedback data, as well as between incorrect and no feedback data. This underscores the evidence that there is a difference between incorrect and no feedback. The
difference between correct and incorrect feedback is also quite evident. There is more of a gray area with correct vs no feedback, which was to be expected given previous research that has stated that feedback will only improve performance if paired with a specific, hard goal.

Many previous studies have explored feedback as a tool to improve this task performance, concluding that feedback improves task performance, but only when coupled with a task performance goal. However, there is no clear consensus on how feedback without goalsetting impacts task performance. This paper aims to investigate feedback’s impact on task performance from a new angle by determining how incorrect feedback affects task performance. Through this investigation, we explored the participant thought process that plays into how the participant uses feedback. This has several implications, for example, in regard to malingering, in which people purposefully provide incorrect feedback for personal gain. To conduct this study, we instructed participants to complete simultaneous tactile discrimination tasks on which they received correct or incorrect feedback. Our results showed a significant difference in task performance between correct and incorrect feedback, and this difference increased for a harder task. Additionally, our results demonstrated that several participants receiving incorrect feedback had longer response times, indicating that they realized they were receiving incorrect feedback. Overall, this study sheds light on how feedback can be used to manipulate task performance.

References