Practice Effects Using Stroop

Renee Romanowski

Indiana University

April 8th, 2016
Abstract

My study focuses on the effect of practice, specifically if starting with the interference condition will make a participant have quicker response times on the neutral condition in a Stroop experiment. Davidson and his team have backed this with data proving that practice makes a difference (Davidson, Zacks & Williams, 2003). I collected data from 16 participants, half of which started with the inference condition PowerPoint first. The participants named a photo while a similar word (for the interference condition) or a neutral word (for the neutral condition) hovered on top. I found that there was no main effects, but there was a strong interaction effect. This is important to find the significance of practice and order in Stroop studies. It is important to take into account that my Stroop test did not have a main effect which could compromise the data.
Practice Effects Using Stroop

Researchers are trained to counterbalance, but that could be a major issue. Specifically in Stroop tests, researchers say it is important to counterbalance to control for the practice effects (MacLeod, 1991). This counterbalancing may be having a major effect on the results of the hundreds to thousands of experiments that have used the Stroop study because the practice effect could pose a bigger issue than previously thought. Instead of controlling for practice effects, researchers might benefit from specifically studying practice effects. Instead of treating these effects as noise and trying to ignore them, they could subject them to experimental study. For example, if a researcher was studying the correlation of Stroop speed and intelligence, it could be the case that the researcher would have to start everybody on the same condition so they do not receive a practice effect from one condition. This would only be the case if it was only possible for one Stroop condition to prepare the participant for the next one, as opposed to either Stroop condition preparing the participant for the next one. This is what I want to find out. My experiment will consist of two conditions, an interference and a neutral condition PowerPoint. Each of 16 participants will take both, but half of them will be specifically started on the interference condition and the other half will be started on the neutral. I will be looking at the difference in response times for the conditions, the groups (start interference and start neutral), and if there was an interaction effect. I want to find out if the order of the conditions can make a significant difference, and I am going to use multiple studies to help shape mine.

There have been several studies that have focused on gender differences in Stroop related tasks, with opposing results. In 2006, Mekarski and his team performed an experiment to put the past experiments to the test to see if there really was a gender difference (Mekarski, Cutmore & Suboski, 1996). The researchers used 16 participants total and had equal amount men and
women. The results were that it is most likely that gender has an effect on color and word Stroop tests, specifically that women are faster than men. The reason for this gender difference is unknown. It is also important to note that color and word Stroop tests are very similar to Picture Word Inference (PWI) Stroop tests, which is the kind I am going to use, so the gender difference should be the same (Maanen, Rijn, & Borst, 2009). Mekarski and his fellow researchers concluded that the women in their study performed on average 46msec faster than men (Mekarski, Cutmore & Suboski, 1996). I do not expect to find a significant change in response time due to gender because my timer only measures to the nearest second, therefore I will randomly select women and men.

As participants begin the study, I will stress speed and accuracy. This will serve two purposes: to make sure the participants understand the rules and to further help eliminate the gender gap. In a study done in 1992 by Kluge, it is shown that high stress situations nearly eliminate the difference in response times for men and women because women speed up when given the direction to go faster (Kluge, 1992). Kluge is defining high stress as formally dressed experimenters who approach the participants and put heavy emphasis on accuracy and speed (Kluge, 1992). I will be approaching people at random which will already create an uneasy environment, and then I will be stressing speed and accuracy. Between the findings of the Mekarski study and the extra work I am going through to eliminate the response time gap, the gender of my participants should have no effect on the outcome of my experiment (Mekarski, Cutmore & Suboski, 1996).

Gender is going to be accounted for, but I also want to avoid age playing a factor in this. Comalli, Wapner, and Werner performed a test in 1962 to understand the effects of Stroop and age (Comalli, Wapner, & Werner, 1962). Although there were significant changes in childhood
and retirement age, there seemed to be little change after the age of 17. People from 17 to 44 performed extremely similarly in this Stroop study. Therefore, for my study I will make sure that all participants fall within that age range, so age does not have an effect on my results.

I have listed multiple experiments that are going to shape my experiment, but I should mention that my specific test has current research backing it up. In 2007 Davidson and his team set out to find out if practice effects in Stroop tests were apparent in all ages (Davidson, Zacks & Williams, 2003). Practice effects in this study means having a participant practice both the interference and neutral condition multiple times before timing them. This in turn makes the time smaller between the neutral and interference when it is actually timed. I am not as interested in the age results of the Davidson study, but the result that practice causes a decrease in the difference in time between the neutral and interference conditions across all ages is what is shaping my study (Davidson, Zacks & Williams, 2003).

The final study which shaped my between subject groups is my own study. In 2016 I conducted a Stroop experiment and noticed that the participants who started with the interference condition performed faster on the neutral condition than the people who started with the neutral condition. This is because the people who received the interference condition first got good practice to do the neutral condition, thus doing it faster. This is why I am going to specially study it with 16 participants instead of only 8 which is what I did in my last experiment (Romanowski, 2016). I also found a main effect of my Stroop condition and I am expecting to repeat that result due to the nature of the PowerPoints being similar.

I believe giving the interference condition before the neutral condition will help train the participant to bypass the conflicting automatic processes and thus they will go quicker in the neutral condition. The participants who receive the neutral condition first will struggle more in
the neutral condition because they had no practice in dealing with the cognitive dissonance that the interference condition provides. So I expect the neutral condition response time to decrease in the group that starts with the interference condition first, but I expect the interference response time to stay the same across both groups.

First, I want to know if my Stroop tests works and if the interference condition takes longer than the neutral condition within all 16 of my participants. Then I want to see if the eight participants who started with the interference condition completed the neutral condition in less time than the group who started with the neutral condition. First a picture word inference version of the Stroop test is going to be given to both groups. One group of 8 people will receive the neutral condition first and the other the interference. All 16 participants will be compared by the individual response times between the neutral and interference condition for the within subjects Stroop condition. Then the two groups will be compared by their average response times across the conditions to see if there is a between-subjects effect of the order of the PowerPoints. I chose to design my study in this way so that each group is given the same exact test that can easily be compared between the two groups.

I hypothesize that the group that is given the interference condition first will be faster than the group that started with the neutral condition on the neutralcondition PowerPoint. I believe this to be true because the way the slides are set up, the group who got interference first will have to become used to ignoring the word on top of the picture that they must name. Thus when they have to do the neutral condition, they will just perform it quicker than the group who did not receive practice beforehand. However the group that performs neutral first will not have the same amount of practice, and thus will not perform as quick on the neutral
condition. Although I do believe the interference condition response time will be similar among both groups.

I expect a main effect of Stroop because my design is very similar to my previous study that elicited a main effect of the interference condition response time being slower than the neutral condition response time. I expect a main effect between my start interference and start neutral groups because Davidson and his team found a significant result in which practice made a participant have a quicker response time (Davidson, Zacks & Williams, 2003). I am also taking into account my previous study in which the interference response time was similar between both groups, but the neutral time was faster in the group that started with the interference condition (Romanowski, 2016). I will know my main effect of Stroop hypothesis is supported if the difference in response times is greater in the group that begin with the neutral condition. If I have a main effect of my between-subject variable then the group that started with the interference condition will go faster on the neutral condition than the groups that started with the neutral condition. Finally, I will have an interaction effect if my previous study holds true and the interference times between the two groups stays the same but the neutral condition response time takes longer in the group that started with the neutral condition. I can conclude that my hypotheses are supported if my p value turns out to be less than .05 in each of the appropriate resamplers. If it is not supported, then I will consult more studies to make sure mine was planned properly and then retest or possibly be able to conclude that practice effects make that big of an impact on response times.

**Method**

There were 16 participants that decided to participate in the experiment. I asked people in the library if they would be willing to name the pictures on the slides of two different
PowerPoints. I had two conditions, the neutral and the interference. The neutral condition was a PowerPoint that consisted of 20 slides of different fruits and vegetables with a neutral word on top. For example, a picture of an apple with the word book on top. The interference condition PowerPoint consisted of 20 slides of the same fruits/vegetables in a different order with an opposite, but similar, word on top. This was meant to confuse the participant, for example a picture of an apple with the word orange on top. Half of the participants started with the neutral PowerPoint, and the other half started with the interference. The two groups were compared to see if there was a significant effect in response times based on the order of the PowerPoints. The independent variable was the order of the conditions and the dependent variable was the response times for my between subjects portion of the experiment. The independent variable for the within-subject Stroop portion was whether the word on top of the picture was a neutral or interfering word and the dependent variable was also response time. I predicted that there was going to be a main effect of my Stroop condition, the interference condition was going to take longer. I also predicted a main effect in the order in which the PowerPoints were administered, the participants who started with the interference condition would be faster at the neutral condition than the participants who started with the neutral condition. I also believed that there was going to be an interaction effect and the start interference and start neutral groups were going to perform the same on the inference condition, but the start interference group was going to be faster on the neutral condition PowerPoint.

**Participants**

There were 16 total participants that were in my study. 8 of my participants begin with the neutral PowerPoint (Start neutral). The other eight begin with the interference condition power point (Start interference). I set up my laptop in the foyer of the library and asked people if
they would be willing to participate in a study. There was no incentive for the participant to come over and participate in my study. After they participated in my study, the participants were asked two questions. The participants were asked if English was their native language in order to increase the probability that they would know the English names for the fruits and vegetables. The participants were also asked their age to ensure none of them fell out of the 17 to 44 age range. I put participants into groups based on the order I recruited them. Odd numbered participants were put into the start neutral group, and even numbered participants were put into the start interference group.

**Materials**

I used PowerPoint to display each of the 40 stimuli to my participants. I also used the Rehearse Timings function in PowerPoint to time them. I obtained my pictures by doing a google search for the specific fruit or vegetable I needed. I decided on the pictures by choosing only fruits and vegetables with white backgrounds, to ensure that background color did not have an effect. I also used the same size font on every slide to keep it the same across on slides. The neutral condition PowerPoint consisted of 20 slides that had a different fruit or vegetable on each of them. Each slide also contained a neutral word written above the photo. The interference condition PowerPoint had the same photos arranged in a different order. The words written above the pictures in the interference condition PowerPoint were meant to be similar but not exact, in order to confuse the participant. As stated in the participant’s subheading, the only questions I asked my participants pertained to their applicability to the study. I did not ask them any questions to decide what group they went into because that was already pre-decided.
Figure 1 – Neutral Condition First Slide

Figure 2 – Interference Condition First Slide

**Procedure**

My sample of participants was a convenience sample. I recruited participants who were at Wells Library on a Sunday by calling them over to the table I had my laptop set up at. I avoided calling anybody over with headphones or people who were walking in groups of people. Many people refused, but 16 people did come over and participated in my experiment. When they came over, I told them they would be going through a PowerPoint and naming the name of what is pictured. I stressed to all the participants that they needed to be as fast, but as accurate, as possible. I also told them that they would be controlling the spacebar themselves in order to move on from slide to slide.

When the first PowerPoint was done, I just told them to repeat the same process for the second PowerPoint that I pulled up while they waited. Once they were done, I asked them if English as their native language and their age. Once they answered, I thanked them and waited for another person. Within-subjects the independent variable was whether the word on top of the photo was a neutral word or an interfering word and the dependent variable was the response time. My between-subject independent variable was which PowerPoint they received first, and
my dependent was also response time. I used the Rehearse Timings function to measure my dependent variable.

**Results**

There were exactly 16 people who participated in my study. Eight people were in the group that specially started with the neutral condition first. The other eight people were in the other group that started with the interference condition first. Table 1 represents my raw data. N/I represents which condition the participant started off with, N for neutral and I for interference. N will be referred to as start neutral, I will be referred to as start interference. The difference is the response time for the neutral condition minus the response time for the interference condition.

Table 1 - Response Times (in seconds)

<table>
<thead>
<tr>
<th>N/I</th>
<th>Participant</th>
<th>Age</th>
<th>Neutral</th>
<th>Inference</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1</td>
<td>19</td>
<td>26</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>N</td>
<td>3</td>
<td>19</td>
<td>24</td>
<td>20</td>
<td>-4</td>
</tr>
<tr>
<td>I</td>
<td>4</td>
<td>19</td>
<td>29</td>
<td>25</td>
<td>-4</td>
</tr>
<tr>
<td>N</td>
<td>5</td>
<td>22</td>
<td>46</td>
<td>38</td>
<td>-8</td>
</tr>
<tr>
<td>I</td>
<td>6</td>
<td>25</td>
<td>29</td>
<td>36</td>
<td>7</td>
</tr>
<tr>
<td>N</td>
<td>7</td>
<td>18</td>
<td>33</td>
<td>24</td>
<td>-9</td>
</tr>
<tr>
<td>I</td>
<td>8</td>
<td>21</td>
<td>29</td>
<td>32</td>
<td>3</td>
</tr>
<tr>
<td>N</td>
<td>9</td>
<td>20</td>
<td>26</td>
<td>25</td>
<td>-1</td>
</tr>
<tr>
<td>I</td>
<td>10</td>
<td>23</td>
<td>26</td>
<td>38</td>
<td>12</td>
</tr>
<tr>
<td>I</td>
<td>11</td>
<td>19</td>
<td>27</td>
<td>25</td>
<td>-2</td>
</tr>
<tr>
<td>I</td>
<td>12</td>
<td>22</td>
<td>27</td>
<td>26</td>
<td>-1</td>
</tr>
<tr>
<td>N</td>
<td>13</td>
<td>20</td>
<td>31</td>
<td>30</td>
<td>-1</td>
</tr>
<tr>
<td>I</td>
<td>14</td>
<td>20</td>
<td>23</td>
<td>25</td>
<td>2</td>
</tr>
<tr>
<td>N</td>
<td>15</td>
<td>19</td>
<td>22</td>
<td>21</td>
<td>-1</td>
</tr>
<tr>
<td>I</td>
<td>16</td>
<td>19</td>
<td>19</td>
<td>21</td>
<td>2</td>
</tr>
</tbody>
</table>

Averages: 27.4 | 27.3 | -0.1
Resamplers help researchers by producing data that is equivalent to repeating an experiment 10,000 times. A resampler takes in the data that was collected and uses it to create 10,000 statistically similar results. This would be synonymous to me collecting data from 160,000 people. A resampler helps to answer if data from an experiment is significant or not. A resampler can help support/deny significance because a more accurate conclusion can be gotten from the more reliable p value that is calculated by the resampler’s simulated people. Researchers resample to save time and come to stronger conclusions. I used a sampling applet to find significance for my within-subjects and between-subject variables (Hullinger, 2006). I also used a different applet for finding an interaction effect (Hullinger, 2006). Calculating significance for the within-subjects variable consists of calculating the difference in the response times for the neutral and interference conditions for all participants and then running the applet to produce statistically similar data 10,000 times. These 10,000 results are used to calculate the confidence interval and significance. Whereas the between-subjects portion for the applet requires the average of the two response times (neutral and interference) for each participant in the two groups to be in separate columns. Then the resampler can be run for 10,000 trials to help find significance. Which in this case significance would mean that the mean difference between the two groups was not accidental. In both the within and between subjects setting of the resampler, the gray box indicates the confidence interval of the 10,000 resamples. The blue line is zero (in within) or the actual mean (in between), and will be outside of the gray box only if the data is significant. If the p-value is less than .05, then that indicates a significant result.
Figure 3 – Histogram of Collected Data Within-Subjects

Figure 4 – Histogram of Sampled Means Within-Subjects

4559.0 out of 10000 fell on the other side of zero, giving p=0.4559
(The 95% CI = [-2.437, 2.5])

Difference in Response Times (s)
The mean response time for the neutral condition was 27.4 seconds and the mean response time for the interference condition was 27.3. 4,559 out of 10,000 resamples fell on the other side of zero, this corresponds to a p value of .5 which indicates there is no significant difference between the neutral and interference conditions. The 95% confidence interval of -2.4 to 2.5 did include zero which means this is not a statistically significant result.

Figure 5- Histogram of Collected Data Between-Subjects

![Histogram of Collected Data Between-Subjects](image-url)
The mean response time for my start interference group was 26.9 and it was 27.8 for mean start neutral group. 3,919 out of the 10,000 resamples fell on the other side of my mean, which equates to a p value of .4. This is not significant because it does not fall under .05. The 95% confidence interval of -4.8 to 4.8 did include my actual mean difference, indicating that there is not a statistically significant difference between the two groups.
The start interference group had an average of 25.5 seconds on the interference conditions and 25.4 seconds for the neutral condition. The start neutral group has a 26.1 second average response time for the interference conditions and a 29.4 second average response time for the neutral condition. The difference of differences was on the other side of zero approximately zero times. This corresponds to a p value of 0 which is less than .05 which means it is significant. The confidence interval of 2.6 to 10.3 did not include zero which also indicates a statistically significant difference.

**Discussion**

I wanted to know if the order of the Stroop conditions made a significant difference in the response times of the participants. I hypothesized there would be a main effect of my Stroop condition with the interference response time being longer. I also believed there would be
main effect of order and the different in response times would be larger for my start neutral group. Finally I hypothesized an interaction effect; I believed that a participant who started with the interference condition would be faster at the neutral than a participant who started on the neutral condition, but a participant who started with neutral would not perform any quicker on the interference condition. I did this by administering a neutral and interference condition to 16 participants, half of which started with the interference condition, and analyzing their response times.

I did not find a main effect of my within-subjects variable since there was no significant difference in response times between the neutral condition and the interference condition. I also did not find a main effect of my between-subjects indicating that what condition the participant started off with did not make a difference in the average of their two response times. However, there was a very significant interaction effect. So the condition the participant started with did affect how quickly they performed the second condition.

The lack of a main effect in my within-subjects variable does not support my original hypothesis. My within-subjects variables of neutral vs interference conditions does not match the findings of the Stroop study it was based off of (Stroop, 1935). In previous Stroop experiments, including Stroop’s own experiments, the interference condition is supposed to take significantly longer than the neutral condition. However, that was not the case in my study because I did not find a difference in the conditions. Therefore, my results did not match previous studies or my hypothesis. This would point to an error in the design or application of my study due to the amount of research that currently supports the Stroop effect.

In accordance with the study done by Davidson, I did expect a significant effect of my between subjects variable in which the group who began with the neutral condition had a larger
average response time between the neutral and interference conditions than the group who started with the interference condition (Davidson, Zacks & Williams, 2003). However I did not find a main effect of the between subjects variables of start inference and start neutral. I did not account for the interference response time changing as much as it did between groups. I expected the interference time to stay similar while the neutral response time was higher in the group that started with the neutral condition. The neutral condition response time was smaller in the start interference group, but the start neutral group also had a lower interference response time than the start interference group. This leads to the understanding of the interaction effects.

My significant result for the presence of an interaction effect did indeed support my original hypothesis. My whole thought behind the reasoning for an interaction effect is learning generally improves the time it takes to complete tasks, such as shown in the study done by Davidson and his Team (Davidson, Zacks & Williams, 2003). I also expected an interaction effect because I had one in my previous study (Romanowski, 2016). However the interaction effect in this study was a bit different than my last. In my last study, when participants had the neutral condition first, their response time in the neutral Stroop was slower than the start interference group and the interference condition response time was the same as the start interference group. However in this study when the participants had the neutral condition first, their response times in the neutral Stroop were slower and in the interference Stroop were faster than the other group. This means it is important to note which participant started with which group because it make a very big difference.

There were a few obvious confounds in this study. When I was switching PowerPoints in front of the participants, they got a good look at the upcoming slides of the second PowerPoint. This possibly made their second time faster. Due to the fact that I let the participant press the
space bar, the participants would sometimes accidently go on even though they named the fruit or vegetable wrong. For example, most participants missed the first two slides on the interference condition because they were too similar, and only realized they were wrong when they got to the third slide and said “orange” when it was visually an apple. The last huge confound is participants not knowing the names of all the fruits and vegetables. I had very little problems with practice participants knowing the names of the food, but quite a few of my 16 participants did not know at least one of the fruits or vegetables. This all affected the response times, possibly making the interference response time faster than it should have been. If I ran the experiment again, I would most likely switch to more common fruits and vegetables and control the spacebar myself.

I think to completely answer my initial hypothesis I would need a working within-subject main effect. Although I got an interaction effect it was different from what I expected. I am very curious as to what would have happened if the participant could not physically go on to the next side until they named the current slide correctly. The first study I would do of this nature would involve making a working within-subject design and then repeating the rest of the experiment. I feel like a main effect of within-subject design would have made this interaction effect more believable. I think another idea to just look at practice effects on Stroop is to do this exact experiment, however give one group each PowerPoint twice, but only time them on the second set. So one group would take the interference and neutral condition PowerPoints one time. The other group would take the interference and neutral condition PowerPoints once for practice, and then a second time to get timed. Then the participants would also be counterbalanced within their groups.
Based on my results, I can conclude that order of conditions in Stroop design does matter in the outcome of the response times. I believed that the start interference group would perform better on the neutral condition than the group that started on the neutral condition, and they did. I believe my conclusion supports the need for counterbalancing. Although I also feel that sometimes the effects of the order of conditions have such an impact on results that they cannot be simply ignored, averaged, or controlled through counterbalancing, but should be studied more. I believe society can take away the knowledge that practice is important and does indeed make you better and faster at things. The results matter because they show practice makes a difference and counterbalancing is important.
References


Hullinger, R. (2006). Running the Sampling Applet. Unpublished manuscript, Department of Psychology, Indiana University, Bloomington, America


Romanowski, R. (2016). Within-Subject Stroop Test Using Picture Word Inference. Unpublished manuscript, Department of Psychology, Indiana University, Bloomington, America