

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/333786424>

Replicating the Stroop Effect

Preprint · May 2019

DOI: 10.13140/RG.2.2.16499.37926

CITATIONS

0

READS

4,233

1 author:



Gabriela Gomez

Nova Southeastern University

1 PUBLICATION 0 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Stroop Effect [View project](#)

Replicating the Stroop Effect

Gabriela Gomez

Florida Atlantic University

Abstract

Process automaticity, interference, and inhibition were studied in this experiment. Specifically, this experiment looked at the automaticity of reading a word versus the not automatic process of identifying color. This experiment utilized congruent and incongruent trials to study the relationship between two different tasks, and to identify whether automaticity interferes with processing ink color. The average speed and accuracy at which 83 participants conducted tasks would illustrate how automatic processing effects the processing of different streams of information. The findings were that on average: 1) congruent trials were completed faster and more accurately than incongruent trials for both levels of task; 2) reading the word occurred faster and more accurately than identifying the ink-color for both levels of congruence; 3) levels of incongruency and task interacted. This research is of great significance as it studies the brain's ability to prioritize automatic processes when clashing with other information.

Keywords: stroop task, automaticity, interference

Replicating the Stroop Effect

In 1935, J. Ridley Stroop published a hallmark paper in the *Journal of Experimental Psychology* titled “Studies of Interference in Serial Verbal Reactions”. Stroop was concerned with how certain mental processes appeared to function involuntarily or automatically and how these automatic processes could interfere with other information processing. Specifically, he studied how a color-word (“red”, “blue”, etc.) could interfere with the task of naming a correlated color or vice versa. In this design, reading was the automatic process being tested. To do so, he used color-word and ink-color pairs composed of a color-word printed in an ink-color that did not correspond (eg: “red” written in blue ink) (Stroop, 1935). Apart from a phase that allowed the participants to get used to the testing format and procedure and gave the researchers baseline data of their participant’s abilities, there were two main phases in Stroop’s experimentation.

These two experimental phases were: 1) interfering color stimuli upon reading names of colors serially, and 2) interfering word stimuli upon naming colors serially. In the first phase, Stroop’s team measured how long it took participants to read aloud a list of color-words printed in ink colors that are different from the color the word describes compared to reading a list of color-words printed in black. In the second phase, participants’ reaction times when naming the ink-color the word was printed in (where the ink-color did not correlate to the color-word) were compared to their reaction times when naming the ink-color of a square block.

Stroop found that it took 2.3 seconds longer to name 100 color-words printed in ink-colors that did not match with each other (eg: “RED” in blue ink), compared to time it took to name 100 color-words printed in black ink. Those few seconds were not considered by Stroop

to be reliable or statistically significant. Also, it took participants significantly longer to name the ink-color of those 100 color-words compared to their reaction times in naming the color-word. Stroop's conclusion was that reading the word was the automatic process and that it interfered with naming the ink-color of the word (which was not automatic). This paper cemented the concepts of automaticity and its interference, and Stroop's findings were coined the 'Stroop effect'.

Hintzman et al's paper (1972) further expanded on Stroop's, illustrating the concept of interference. In their experiment, they focused on two possible explanations for the Stroop effect. This effect could have been explained by the color-word and ink-color working as stimuli that trigger two separate pathways in the brain running parallel to each other and therefore, competing for 'attention'. Alternatively, the effect could be that the color-word interferes or distracts the brain from processing the ink-color (Hintzman et al, 1972). In other words, did interference occur when inputting information to process or when "deciding" which information to output?

Hintzman et al (1972) utilized a measure of reaction times for color-naming under four conditions: 1) congruent trials where color-word matched the ink-color; 2) incongruent trials where color-words and ink-colors did not match; 3) neutral words matched with random ink-colors; and 4) colored non-words. The experiment's results (mainly that there were faster reaction times with respect to the congruent condition compared to color-naming non-words) indicated that making the color-word and ink-color match facilitated the processing of the information required for that trial. And therefore, it supports the theory that the brain's processes

for reading and color-naming run in parallel and compete with each other for dominance in the tasks. Hintzman et al assert that the Stroop effect is an ‘output’, not ‘input’ based phenomenon.

With respect to the Stroop effect, there has been some debate in recent years as to experimental interventions that could decrease or remove the interference caused by automatic processes and the actual mechanism behind automatic processing. Kinoshita’s, De Wit’s, and Morris’ (2017) work on reading automaticity illustrates this argument. They used five distinct types of “distractors” apart from the standard incongruent trials. They specifically tested the response time when a participant was asked to name the ink-color out loud versus when the participant indicated the ink-color by pressing on a key. They found that when naming the ink-color aloud, the distractors interfered at different levels. Some were very distracting in the task, and some were less distracting. When pressing a key to indicate the ink-color however, all five types of distractors interfered equally. This led the experimenters to conclude that automatic reading is not an unchanging property driven by stimulus (lexicality of a word). Rather, automaticity is controlled by the goal of the task: to speak aloud or to indicate manually.

There is much research already done on the topic and evidence that is well accepted by the community. Such recent discussion on an effect that was first recorded in 1935 indicates that there are still questions left unanswered with respect to the mechanisms behind Stroop effect. I would support the legitimacy of this experiment by bringing attention to the field of psychology’s current replication crisis.

The first collaborative replication project was Open Science Collaboration in 2015. Out of the 100 cognitive and social psychology studies they intended to replicate, 97 had significant findings. However, they could only replicate the results of 36 of those studies (Stroebe, 2019).

Similarly, Many Labs was a project group composed of 186 researchers from around the world. Their work consisted of repeating (or attempting to repeat) a set of 28 previously printed findings. Despite the sheer number of labs and sample sizes used to replicate the results of these 28 studies, only 14 were replicable. The 14 experiments that were replicated varied in their agreement with the original findings (Klein, 2018). This was one of several projects dedicated to replicating accepted findings that has concluded with similar results. All together, these studies further demonstrate the scientific importance of experiments being replicable before the results can be accepted by the general scientific community.

In replicating Stroop's experiment, it is my hope that the results will be at the very least, in line with Stroop's original findings. My hypothesis is in four parts: 1) the mean reaction time for the congruent reading task will be faster than the congruent ink-color identification task; 2) the mean reaction time for the incongruent reading task will be faster than the incongruent ink-color identification task; 3) the mean accuracy rate for the congruent reading task will be higher than for the congruent ink-color identification task; and 4) the mean accuracy rate for the incongruent reading task will be higher than for the incongruent ink-color identification task. Overall, I predict that congruent tasks will be completed faster and more accurately than incongruent tasks, and that word-reading tasks will also be completed faster and with higher accuracy than ink-color identification tasks.

METHOD

Participants

There were a total of 83 participants (24 male, 59 female) that were all students in an undergraduate level research methods course at Florida Atlantic University. Participants were

compensated with course credit by participating in the experiment. The average age of the participants was 21.52 years ($SD = 2.71$), with the youngest participant being 17 and the oldest being 37. Thirteen participants were left handed, the other 70 were right handed. The author (GG) was also a participant as well as a researcher. Participants were treated in accordance with the “Ethical Principles of Psychologists and Code of Conduct” (American Psychological Association, 1992).

Materials

The experiment was run on computers in a computer lab on the Florida Atlantic University campus. The program used to run the experiment was E-prime3 software, and participants used the keyboard to indicate their answers. An experimenter gave the participants instructions in the computer lab.

Each participant was randomly assigned to one of two groups and given an individual ID number. This was both a between-subjects and within-subjects design. As a between-subjects design, groups were created to balance the order of the tasks conducted, and participants scores in each task and level of congruence were compared to each other within groups. This was a 2x2x2 factorial design with three independent variables. First, there was the level of congruency in each task, congruent or incongruent. Second, there was the task itself, either reading the word and ignoring the ink-color, or ignoring the word and identifying the ink-color. Finally, there was the order the tasks were given in: identifying the word, then identifying the ink-color or identifying the ink color, then identifying the word. There were two dependent variables being measured: the reaction time per list and the overall accuracy for each task.

There were three tasks in this experiment, based on the list of color-words and ink-colors from the original Stroop experiment (1935). The list in this experiment was slightly altered. Only three colors were used: green, blue, and red. Each participant completed one practice task, one word-reading task, and one color-identifying task. Each of these tasks consisted of list items that were congruent or incongruent.

Both groups completed all three tasks, but in different orders to battle possible order effects in the experiment. Both groups completed the practice task first, but the experimental tasks were given in different orders. Group 1 conducted "Stroop cw" first. This task required the participants to press the key corresponding to the color-word that appeared on the screen. The point of this task was for participants to read the word. Group 2 conducted "Stroop wc" first. This task required the participants to press the key corresponding to the ink-color of the word that appeared on screen and was conducted with the purpose of having the participants ignore the reading of the word and identify the ink-color instead.

All three tasks consisted of 45 trials; the experimental trials consisted of an approximate mix of 15 congruent and 30 incongruent trials presented one at a time on the screen. There were three color-words and the same three ink-colors (green, blue, red). In congruent trials the color-word and the ink-color matched (the word "green" in green ink). In incongruent trials, they did not match (the word "green" in blue ink). In the practice task, all 45 trials consisted of the color-words written in black ink. Participants were measured for accuracy and for reaction time, with the emphasis being on the overall accuracy and reaction time for each task. At the end of each trial, there appeared a feedback screen that informed the participant of their accuracy and reaction time in the trial, and their overall accuracy in the task so far.

Design and Procedure

Participants were instructed by the proctor to open up the E-prime3 program on their computer and load and run the "Stroop bw" task. This task was a practice task which both groups conducted first. In this task, participants were acquainted with the program and the experimental procedure. Participants first had to answer several prompts that asked for their Group # and ID. Participants were instructed to place their left ring finger on the "1" key, their middle finger on the "2" key, and their pointer finger on the "3" key on the keyboard. At the bottom of the screen, there was a legend: "1" for green, "2" for blue, "3" for red. These instructions would be used for all three tasks. Participants practiced pressing the keys that corresponded to the word that appeared on screen. They were measured based on accuracy and on reaction time, however the proctor instructed participants to prioritize accuracy over reaction time.

Next, the experimental tasks were conducted. Participants had to answer the same prompts, had the same legend at the bottom of the screen, and utilized the same three keys in the same manner as in the practice task. Participants pressed the keys corresponding to the new instructions for each task (to identify the color-word or to identify the ink-color). Again, the proctor instructed the participants to concentrate on accuracy and not to dwell on the information presented on the feedback screen.

RESULTS

An alpha level of $p < 0.05$ was used to denote the statistical significance of the results. The independent variables tested were task and congruence, with reaction time and accuracy being the dependent variables. Dependent variables were measured in milliseconds and average rate. The order of the task was controlled for by counterbalancing the order with which tasks

were given to different groups of participants. There were no significant order effects either on reaction time or accuracy.

With respect to reaction time, a 2x2 analysis of variance (ANOVA) was conducted to see if reaction time was dependent on the conditions of congruency and of task, as well as to identify any possible interaction between congruency and task. There was a significant main effect of congruence on reaction time as $F(1, 81) = 171.56, p < .001, \eta p^2 = .68$. With $\eta p^2 = .68$, congruence had a large effect on reaction time. The mean response time for the congruent condition was significantly faster ($M = 657.57, SD = 15.23$) than for the incongruent condition ($M = 756.31, SD = 17.62$) Refer to Table 1.

There was a significant main effect of task on reaction time as well as $F(1, 81) = 21.42, p < .001, \eta p^2 = .21$. The effect size of task on reaction time was medium in size. It took participants less time in naming the word ($M = 672.38, SD = 18.41$) compared to the ink color task ($M = 741.50, SD = 17.15$) Refer to Table 2.

Also, there was a significant interaction between congruence and task on reaction time as $F(1, 81) = 24.18, p < .001, \eta p^2 = .23$ with the effect size being medium. The reaction time of the participants was dependent on both the condition of congruency and the condition of the task.

A paired-samples t-test was performed to test the hypothesis that the mean reaction time for the congruent task of reading the word was faster than the mean reaction time for the congruent task of identifying the ink color. The test results were $t(82) = 1.96$ and $p = 0.053$. As the p-value is higher than the alpha level of 0.05, this is not significant enough to make a conclusion with respect to level of task and congruence. Mean reaction times for the congruent

word naming task ($M = 641.62$, $SD = 17.50$) were not significantly faster than mean reaction times for the congruent ink color identification task ($M = 673.53$, $SD = 16.78$).

Another paired-samples t-test was conducted to test the hypothesis that the mean reaction time for the incongruent word reading task was faster than the mean reaction time for the incongruent color identification task. The test was significant as $t(82) = 6.00$ and $p < .001$. This showed that mean reaction times for the incongruent word naming task ($M = 703.15$, $SD = 18.41$) were indeed significantly faster than mean reaction times for the incongruent ink color identification task ($M = 809.47$, $SD = 18.91$).

With respect to accuracy, another 2x2 ANOVAs were performed to see if accuracy was dependant on the conditions of congruency and of task, as well as to identify any possible interaction between congruency and task on accuracy. There was found to be a significant main effect of congruence on accuracy as $F(1, 81) = 58.58$, $p < .001$, $\eta p^2 = .42$. Congruence was found to have had a large effect on reaction time. The mean accuracy for the congruent condition was significantly higher ($M = 98.80$, $SD = 0.30$) than for the incongruent condition ($M = 94.30$, $SD = 0.50$) Refer to Table 1.

There was a significant main effect of task on accuracy as well as $F(1, 81) = 4.51$, $p = 0.037$, $\eta p^2 = 0.05$. The effect of task on accuracy was small in size. Participants were more accurate in naming the word ($M = 97.10$, $SD = 0.041$) compared to the ink color task ($M = 96.00$, $SD = 0.40$) Refer to Table 2.

Finally, there was a significant interaction between congruence and task on accuracy as $F(1, 81) = 8.88$, $p = 0.004$, $\eta p^2 = 0.10$ with the effect size being medium. The accuracy of the participants was dependent on both the condition of congruency and the condition of the task.

A paired-samples t-test was performed to test the hypothesis that the mean rate of accuracy for the congruent task of reading the word was higher than the mean rate of accuracy for the congruent task of identifying the ink color. The test results were $t(82) = 0.86$ and $p = 0.39$. As the p-value is 0.39 and much larger than the established alpha level of 0.05, the null hypothesis can not be rejected. This showed that mean accuracy for the congruent word naming task ($M = 98.60$, $SD = 0.49$) was not significantly higher than mean accuracy for the congruent ink color identification task ($M = 98.90$, $SD = 0.30$).

One last paired-samples t-test was conducted to test the hypothesis that the mean accuracy rate for the incongruent word reading task was higher than the mean accuracy rate for the incongruent color identification task. The test was significant as $t(82) = 2.84$ and $p = 0.006$. This showed that mean accuracy for the incongruent word naming task ($M = 95.50$, $SD = 0.60$) was in fact higher than mean accuracy for the incongruent ink color identification task ($M = 93.00$, $SD = 0.60$).

Discussion

The purpose of this experiment was to re-test Stroop's original automaticity experiment with respect to the replication crisis in psychology. The experimental findings in question were related to automaticity and interference (Stroop, 1935). In short, automatic processes (reading) should take superiority in the brain when a second task (color identification) is given. In this experiment, this would have been illustrated with two levels of task: naming the word and identifying the ink-color. Furthermore, there is also a focus on automatic processes "working together" with other process to analyze information better (Augustinova & Ferrand, 2014). This

aspect of the experiment was carried out using another variable, congruency with two experimental levels, congruent and incongruent.

My hypothesis was therefore, two-fold; that congruent trials would be faster and more accurate than in incongruent trials, and that reading the word would be a faster and more accurate task than naming the ink color due to automatic processing. When dividing my hypothesis by levels of the independent variables, I made four predictions: 1) the mean reaction time for the congruent reading task will be faster than the congruent ink-color identification task; 2) the mean reaction time for the incongruent reading task will be faster than the incongruent ink-color identification task; 3) the mean accuracy rate for the congruent reading task will be higher than for the congruent ink-color identification task; and 4) the mean accuracy rate for the incongruent reading task will be higher than for the incongruent ink-color identification task. Trials were measured in reaction time and accuracy. This experiment found all four hypothesis to be true and therefore support Stroop's original findings.

The results for response time and accuracy were looked at separately and each was organized for reading the word and naming the ink under congruent and incongruent trials. This way measurements could be compared across congruency levels and task type together. With respect to reaction time, congruence had a large effect, task had a medium effect, and the interaction of congruence with task type also had a medium effect. In terms of accuracy, congruence again had a large effect, task had a small effect, and the interaction effect was still medium. The fastest and most accurate trials were congruent compared to incongruent. The fastest reaction times were in reading the word tasks versus in identifying the ink-color. The

same pattern appears with accuracy, but to a much lesser degree. The data support the hypothesis. In turn, the data also reflects Stroop's original findings (1935).

Regarding the theoretical implications of this research, the process of automaticity is experimentally sound in its role as the prevailing processing mechanism (Augustinova & Ferrand, 2014). Reading is an automatic process because through practice, the brain is capable of processing the information requiring little active attention and little effort to do so (Walczyk, 2000). The automaticity of reading should have facilitated ink-color identification when the trial was congruent. Whereas in incongruent trials, the automatic reading process should have interfered with color identification (Hintzman et al., 1972). This experiment's results support both assertions.

Congruent trials were on average faster and more accurate than incongruent trials for both levels of task, illustrating that with a lack of conflict, automaticity does not interfere. Kinoshita et al, further claim that with lack of conflict, automaticity can facilitate as the streams of information coincide (are congruent) by working together with the non-automatic information to speed up processing overall. (Kinoshita et al., 2017). This experiment focused on inhibition in incongruent conditions, but facilitation in congruent conditions could potentially be just as significant. With respect to our data, reaction times were faster in congruent tasks compared to incongruent ones. ($MM = 657.57 < MM = 756.31$). Congruent tasks were also more accurate than incongruent tasks ($MM = 98.80 > MM = 94.30$). This experiment's results do not directly support Kinoshita et al.'s work, but further study may be needed to come to a conclusion either way.

Concerning the validity of the experiment, the effect size of each variable and their interaction on reaction time and accuracy was analyzed. The statistical significance of the findings was well grounded. It is necessary to note that the concepts being studied are processing mechanisms that are being quantified with reaction time and accuracy. These two dependent variables illustrated the relationship between automatic processing, facilitation, and interference. Slower times and lesser accuracy corresponded to interference; faster times and greater accuracy to facilitation.

There were two possible confounds that I could identify. The first was the use of a key at the bottom of the screen during the experiment. This key (1-green, 2-blue, 3-red) was present not only in the practice trial, but through all experimental trials. It is a possibility that the presence of this key may have further facilitated in the reading task and/or further interfered in the color identification task as participants may have found it simpler to “match” the list words to the ones at the bottom of the screen. This may have skewed the results. I propose to make a key where “1” is in green ink, “2” is in blue ink, and “3” is in red ink instead. Second, part of the instructions was to emphasize accuracy to the participants. When participants were more focused on accuracy than reaction time, the reaction times represented the effect of the independent variables and their interactions better than accuracy rates. It may be that if participants were instructed to focus on reaction time over accuracy, there may be different results. This is worth looking into in future experiments.

The cognitive mechanisms at play here are mainly automaticity and interference. This study found that on average reading a word occurs faster than identifying the color of the word, illustrating the dominance of automatic reading processes over the non-automatic color

identification processes. Also, on average, congruent color-identification trials were faster and more accurate than incongruent color-identification trials, showing that in trials where the two streams of information did not correspond (eg: “blue” in red ink), automatic reading processes interfered in analyzing the information. In the trials where the two streams of information did not correspond (eg: “blue” in red ink), the automaticity of reading interfered in analyzing the information from ink-color.

References

- Augustinova, M., & Ferrand, L. (2014). Automaticity of Word Reading: Evidence From the Semantic Stroop Paradigm. *Current Directions in Psychological Science*, 23(5), 343–348. <https://doi.org/10.1177/0963721414540169>
- Hintzman, D. L., Carre, F. A., Eskridge, V. L., Owens, A. M., Schaff, S. S., & Sparks, M. E. (1972). "Stroop" Effect: Input or Output Phenomenon? *Journal of Experimental Psychology*, 95(2), 458-459.
- Kinoshita, S., De Wit, B., & Norris, D. (2017). The magic of words reconsidered: Investigating the automaticity of reading color-neutral words in the Stroop task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(3), 369-384.

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Jr., Alper, S., ... Nosek, B. A. (2019, February 11). Many Labs 2: Investigating Variation in Replicability Across Sample and Setting. *Advances in Methods and Practices in Psychological Science*, 1(4), 443-490. <https://doi.org/10.17605/OSF.IO/8CD4R>

Stroebe, W. (2019). What can we learn from many labs replications? *Basic and Applied Social Psychology*, 41(2), 91-103. <http://dx.doi.org.ezproxy.fau.edu/10.1080/01973533.2019.1577736>

Stroop, J. R. (1935). Studies of Interference in Serial Verbal Reactions. *Journal of Experimental Psychology*, 18, 643-662.

Walczyk, J. J. (2000), The Interplay Between Automatic and Control Processes in Reading. *Reading Research Quarterly*, 35: 554-566. <https://doi.org/10.1598/RRQ.35.4.7>

Table 1.

Reaction Time (in ms) for Levels of Task and Congruency

		Task		
Congruency		Identifying Color	Naming Word	Marginal Mean
Congruent	<i>M</i>	673.53	641.62	657.57
	<i>SE</i>	16.78	17.50	15.23
Incongruent	<i>M</i>	809.47	703.15	756.31
	<i>SE</i>	18.91	20.84	17.62
Marginal Mean	<i>M</i>	741.50	672.38	
	<i>SE</i>	17.15	18.41	

Table 2.

Accuracy Rate for Levels of Task and Congruency

		Task
--	--	------

Congruency		Identifying Color	Naming Word	Marginal Mean
Congruent	<i>M</i>	98.90	98.60	98.80
	<i>SE</i>	0.30	0.49	0.30
Incongruent	<i>M</i>	93.00	95.50	94.30
	<i>SE</i>	0.60	0.60	0.50
Marginal Mean	<i>M</i>	96.00	97.10	
	<i>SE</i>	0.40	0.41	