

The Effect of Language Proficiency and Automatic Linguistic Correcting on Interfering Intensity

Dan Geng^{1, *, †}, Yixiang Liu^{2, †}, Jingkang Wang^{3, †}

¹Department of Counseling and Human Development, George Washington University, DC, United States

²Department of Psychology, University of Washington, Seattle, United States

³Department of Psychology, McGill University, Montreal, Canada

*Corresponding author. Email: sylvia4dan@gwu.edu

†These authors contributed equally.

ABSTRACT

The Stroop Color-Word Test (SCWT) is a psychological test used to assess the semantic interference that is manifested as the difference in reaction time between processing congruent (color-meaning matched) and incongruent (color-meaning mismatched) word conditions, and has been well studied in the cognitive field across different countries. However, further studies on factors that affect such interference still lack consistencies, especially in terms of the influence of language proficiency among Chinese-English bilinguals. Additionally, an unconscious mental process that is hard to inhibit is also pervasive within language comprehension itself. Usually, people can recognize word meanings without thinking seriously or spend cognitive resources to decode words, and can even automatically correct a string of words filled with typos or perversions and successfully get the right semantic information due to top-down processing. To explore the influence of these two linguistic factors on the Stroop effect, Chinese-English bilinguals (native Chinese) were recruited as participants online, and their response time and accuracy in traditional SCWTs in English and Chinese version, as well as a modified English version SCWT which contains pseudowords resembling color words, were recorded and analyzed for comparison. The study found a significant Stroop effect within each language version among participants, but no effect of language proficiency on semantic interference, as there was no response time difference between tests in two languages. Moreover, the influence of automatic linguistic correcting was found in SCWT, supported by a significant congruency effect and facilitating effect in English pseudowords.

Keywords: Stroop Color-Word Test, Chinese-English Bilingual, Psycholinguistic.

1. INTRODUCTION

Stroop Color-Word Test (SCWT) has been well studied across the cognitive research field since it was first discovered in 1935 by Stroop. The original test was designed to examine the interference in the recognition of two types of information simultaneously by asking participants to react to different color-word strings. By showing participants different types of color-word (words' meaning congruent or incongruent to the ink colors), the experiment showed a significant difference in response time between processing color stimulus and word stimulus, and the reaction time difference between incongruent groups and congruent groups accounts for the semantic interference (i.e., Stroop effect or Congruency effect) [1].

Nearly a century after the Stroop effect was discovered and illustrated, a profusion of studies has been conducted by psychological researchers to study subjects related to selective attention, cognitive flexibility, processing speed, and inhibitory control [2-5]. Additionally, various versions of the Stroop test were developed in different language scenarios, and the Chinese version is among them, which is considered as a valid method to test cognitive inhibition with high reliability [6-8]. Nonetheless, a more specific group of bilingual individuals, native Chinese speakers who also speak English fluently, have not been thoroughly studied. Moreover, some of the few studies involving this group of Chinese-English bilinguals as participants [9], experienced replication crisis and yielded controversial results which were inconsistent with theories from previous literature. As Magiste has

proposed and corroborated in her experiment, for bilinguals, if the word stimulus is presented in the dominant language (the language that the participants are more proficient with), it will have a larger Stroop interference effect despite the response languages they use [10]. However, Chen and Ho suggested that besides language proficiency, other factors in the experiment may also affect the intensity of interference in Stroop tests in different languages, accounting for the inconsistencies with Magiste's study [9]. To fully explore this topic and settle this dispute on the relationship between language proficiency and the Stroop effect, more relevant research is needed.

Furthermore, the unconscious mental process which is hard to inhibit is also pervasive within language comprehension itself. Previous studies have suggested that automatic processing of word reading is essential for understanding word meanings: People can recognize word meanings without thinking about them seriously and do not have to spend cognitive resources to decode words [11]. Sometimes readers even can automatically correct a string of words filled with typos and successfully process the correct semantic information due to top-down processing. However, the influence of such reading effects in Stroop tests still lacks investigation. Thus, to have a better understanding of the automatic linguistic process, it is also of great significance to explore the effect of automatic linguistic in Stroop tests by introducing 'pseudowords' (words with wrong spelling) that resemble color words, to see whether similar semantic interference remain.

In summary, there are two aims and corresponding hypotheses of the present study. The first one is to test whether the linguistic interference in the Stroop test in English may be influenced by participants' relative proficiency in speaking English compared with their native language, Chinese. Therefore, we hypothesize that comparatively speaking, the semantic interference will be stronger in Chinese for English-Chinese bilinguals who speak Chinese as their first language, since they are more proficient and familiar with their native language. The second one is to test whether the interference in the Stroop test still holds under the circumstance when color words are replaced by non-words that look similar to the actual color words since there might exist automatic linguistic correcting where people tend to automatically correct them and process the meaning of the actual words that are familiar to them. We hypothesize these English 'pseudowords' can also generate congruency effect and interference effect, which is similar to original real color words in incongruent conditions.

2. METHODS

2.1. Participants

In this study, twenty-one participants were recruited randomly online through experiment links of Pavlovia platform. Participants were screened by their self-report personal information and all of them had no color blindness. They (N = 21) were Chinese bilinguals who speak Chinese as their primary language and speak English as a second language. Two participants were excluded due to unqualified data. Finally, data from 19 participants (6 males and 13 females) were analyzed. All the participants are more familiar with Chinese and have a relatively high proficiency and preference compared with English. The sample consisted of participants ranging in age from 20 to 28 years old (Mage=21.84; SD=2.29). This study has been conducted following the statement of ethical principles according to the responsible committee on human experimentation and informed consents were obtained from all participants.

2.2. Materials

Stroop Color-Word Test (SCWT). Three groups of color-word lists were included in this Stroop Color-Word Test (Chinese version, English version, and English pseudowords version) and 66 items in total. Chinese version and English version presented three colors (red, blue, and orange) making a total of 18 word-color pairs. One list was the Chinese version which used Chinese characters and included three conditions: six congruent items (color words presented inconsistent color), six incongruent items (color words presented in inconsistent color), and six neutral items (words that do not have any color meanings). The other list was the English version which presented English words and included three conditions: six congruent items, six incongruent items, and six neutral items as well. The third list was a pseudoword version based on the English Stroop test where items were presented in three colors (yellow, green, and purple). Pseudoword items are defined as meaningless non-words resembling color words and were constructed by exchanging two adjacent letters in the original color words in our test (i.e., yellow to 'yelolw', green to 'geren', and purple to 'purpel'). Thirty items included five conditions: six congruent items, six incongruent items, six neutral items, six pseudo-congruent items (e.g., 'yelolw', the pseudoword of yellow in yellow ink), and six pseudo-incongruent items (e.g., green is written in 'geren' and is presented in purple).

2.3. Procedures

Participants clicked the experiment link online and used computers to complete this test. First, participants

rated from one to ten to self-report their English and Chinese proficiency. Then, they completed three groups of computer-based Stroop tests which were designed by PsychoPy. To avoid the effect of order when testing language effect, which of the language version of the Stroop test appeared first was random. Throughout the whole task, words were showed at the center of the screen, and participants were required to press corresponding keys to indicate the color of the stimuli as fast and accurately as possible regardless of the meaning. The accuracy and reaction time for each item were recorded.

English Version and Chinese Version Stroop Test.

The first two groups were traditional Stroop tests delivered in Chinese and English. Each language version of the Stroop test contained 18 items, all of which were displayed in random order. A fixation cross was displayed in the center of the screen for 700ms and followed by a color word or a neutral word with congruent or incongruent color which had no time limit.

English Pseudowords Stroop Test. This test was based on the English Stroop test with the addition of 18 pseudowords. Thirty items in total were randomly displayed in this trial. Participants were displayed with a fixation cross for 700ms in the center of the screen and followed by a color word, neutral word, or pseudoword with congruent or incongruent color with no time limit as well. Figure 1 is the example flow of this test.

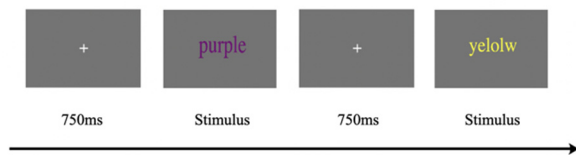


Figure 1. Experiment Flow of English Pseudoword Test

2.4. Statistical Analyses

The reaction time of each condition was collected and computed by Excel and Python. Descriptive statistics including means and standard deviation of reaction time and response accuracy rate were analyzed. Reaction times were analyzed using Z scores and those values smaller or larger than three standard deviations of the mean were discarded in each condition. One participant was discarded due to the low response accuracy of choosing the color. One participant was discarded because of the long response time. The means of reaction time between the Chinese version and English version in each condition were compared. The means of reaction time were compared between pseudowords conditions. ANOVA was used to see if there were statistically significant differences between these means to identify if language proficiency could affect the ability to inhibit cognitive interference and if

there were significant interference effects of pseudowords.

For Aim 1, the experiment used a 2x3 design and was analyzed by two-way ANOVA. Independent variables were two versions (Chinese and English) and three conditions (congruent, incongruent, and neutral). The dependent variables were reaction time and accuracy rate. For Aim 2, the independent variables were word conditions (congruent, incongruent, neutral, pseudo-congruent, and pseudo-incongruent). The dependent variables were accuracy rate and reaction time. One-way ANOVA (including post-hoc tests, if applicable) was used to test the semantic interference in English pseudowords.

3. RESULTS

3.1. Language Effect

For Aim 1, data were analyzed using Excel and Python to report means of response time and accuracy rate, as well as compare these descriptive statistics in different conditions. Firstly, the response accuracy rates of each condition in the first two Stroop tests were displayed in Figure 2, where the incongruent group in the Chinese Stroop test possessed the lowest accuracy rate (88.33%) among all conditions. Obviously, within each language version, it was always the case that the congruent group had the highest accuracy, the neutral group came after, and the incongruent group had the lowest one.

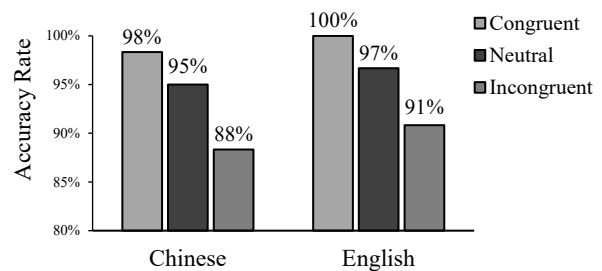


Figure 2. Response Accuracy Rate of Each Condition in Different Language Conditions

The means and standard deviations of response time in all conditions were calculated and the summary of these descriptive statistics is displayed in Table 1. The effect of Stroop test was significant in both English version $F(2, 322) = 4.827, p = .009$ and Chinese version $F(2, 320) = 3.075, p = .048$. The mean response time of incongruent conditions was significantly higher than those of congruent and neutral conditions, within both the English and Chinese versions. The comparison of mean response time in different conditions is illustrated in Table 1. In each condition, the difference between the English version and the Chinese version was conducted as well. There were no significant differences between

English version and Chinese version in congruent condition $F(1, 636) = .016, p = .900$ and incongruent condition $F(1, 626) = 2.458, p = .117$. There were no significant differences between English and Chinese version in neutral condition $F(1, 626) = 1.641, p = .201$ as well. Mean Response Time in Two versions of Stroop test was shown in Figure 3.

Table 1. Mean Reaction Time (ms) and SD of English Version (Eng) and Chinese Version (Chi)

	Congruent (M ± SD)	Incongruent (M ± SD)	Neutral (M ± SD)	F	p
Eng	845.58 ± 432.18	1020.93 ± 476.25	874.30 ± 413.92	4.827	.009
Chi	852.48 ± 368.33	931.22 ± 452.22	801.86 ± 307.26	3.075	.048

Note. The mean difference is significant at the .05 level.

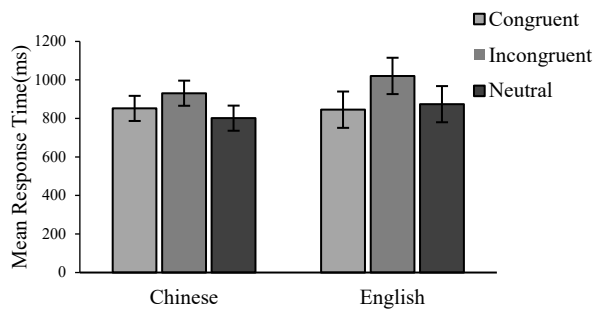


Figure 3. Mean Response Time in Two Versions of Stroop Test

3.2. Pseudoword Effect

For Aim 2, we reported descriptive statistics for different word conditions in the third Stroop test including English pseudowords, as displayed in Table 2. Figure 4 compared the response accuracy rates of each condition in the third Stroop test, showing the outcome that the accuracy rates of congruent and pseudo-congruent conditions were the highest (100%), while the pseudo-incongruent group distinctly had the lowest accuracy rate (91.15%) among all conditions. The Mean response time of different conditions in the pseudoword Stroop test is illustrated in Figure 5.

Table 2. Parameters of Response in Different Word Conditions in the Third Stroop Test

	Con	Incon	Neutral	Pseudo-con	Pseudo-incon
Number of Items	112	112	112	112	113
Mean Response Time (ms)	840.34	1033.52	988.62	779.62	1040.34
Std. Deviation	466.26	477.93	541.39	366.25	500.98
Accuracy Rate	100.00%	95.54%	97.32%	100.00%	91.15%

Note. Accuracy Rate = The number of items participants reacted correctly divided by the number of total items displayed in each condition.

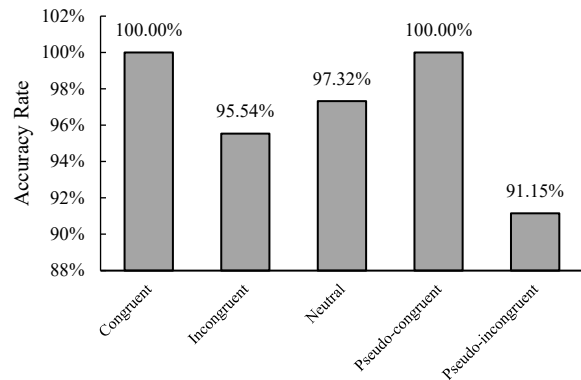


Figure 4. Response Accuracy Rate of Each Condition in the Third Stroop Test

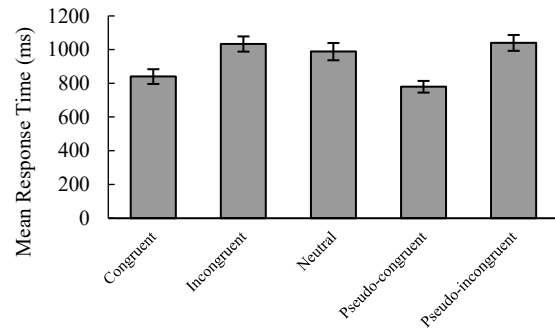


Figure 5. Mean Response Time (ms) of Conditions in the Third Stroop Test

The study used one-way ANOVA analysis to compare response time in different conditions setting a significant level $\alpha = .05$. Since the Levene's Test for Equality of Variances suggested a violation of the assumption of homoscedasticity ($p = .005$), Welch's ANOVA was adopted for homogeneity correction. The analysis revealed a statistically significant effect of word condition on participants' response time, $F(4, 276.547) = 8.483, p < .001$. To understand the significant overall test and further explore semantic interference of interest by investigating pairwise response time difference, we conducted Bonferroni post-hoc comparisons among all pairs of group means (with $\alpha = .05$). Just as demonstrated in Figure 5, the mean response time of incongruent condition ($M = 1033.52, SD = 477.93$) was found to be significantly greater than that of congruent condition ($M = 840.34, SD = 466.26$), $t = -3.048, p = .024, d = -.409$, which was consistent with the congruency effect in traditional Stroop tests. In terms of pseudoword conditions, participants exhibited significantly longer response time processing items in pseudo-incongruent condition ($M = 1040.34, SD = 500.98$) than items in pseudo-congruent

condition ($M = 779.62$, $SD = 366.25$), $t = -4.124$, $p < .001$, $d = -.594$, indicating a significant congruency effect among English pseudowords. In addition, participants took significant shorter time in response to stimuli in pseudo-congruent condition than stimuli in neutral condition ($M = 988.62$, $SD = 541.39$), $t = 3.298$, $p = .010$, $d = .452$. Also, incongruent condition and pseudo-incongruent condition both had longer response time when compared with neutral condition, but neither difference was significant.

4. DISCUSSION

The purposes of the research were investigated. For Chinese-English bilinguals who speak Chinese as their first language, whether the semantic interference of processing word meaning is stronger in their native language. In addition, whether the phenomenon of automatic linguistic correcting takes effect in the Stroop test, in a way that similar congruency effect remains under the circumstance where color words are replaced by non-words that resemble the original color words.

Based on the response time in the Chinese Stroop test and English Stroop test, significant semantic interference was found both within Chinese and English. In addition, reduced accuracy in incongruent conditions for both English and Chinese versions are also compatible with previous literature, serving as evidence of automatic interference that exists pervasively across different languages. However, in terms of the influence of language proficiency, the intensities of semantic interference of Chinese and English were not significantly different. Also, no significant difference in response times in each condition between the Chinese version and English version Stroop test was found.

For the influence of automatic linguistic correcting, we found a significant congruency effect in English pseudowords (response time difference between pseudo-congruent and pseudo-incongruent conditions) and a significant facilitation effect of pseudo-congruent items, along with saliently reduced accuracy in incongruent conditions compared with congruent, pseudo-congruent, and neutral conditions. These results together support the idea that automatic linguistic correcting exists and takes effect in the Stroop test. The corresponding tentative explanation is that people tend to automatically correct English pseudowords and process the meaning of the corresponding real words, which facilitates people's decision by providing semantic clues when the ink color matches the meaning of original words. On the other hand, this effect also makes people react slower and make more errors in color-meaning mismatched conditions, since they need to use additional mental resources to inhibit interference of this automatic linguistic processing (i.e., linguistic correcting and semantic retrieval) when judging the ink color.

Some results are not in line with our assumptions: No influence of language proficiency was found; The response time between the Chinese version and English version is not significantly different; In pseudoword Stroop test, no significant response time difference between incongruent and neutral conditions, nor the difference between pseudo incongruent and neutral conditions was found.

These unexpected outcomes draw forth some non-negligible limitations that potentially undermine the validity of our research. In terms of participants, this study has a relatively smaller sample size, making the outcomes vulnerable to chance factors. Also, due to convenient sampling, our participants share great homogeneity (e.g., similar educational background, narrow age range), which causes difficulty in generalizing our conclusions to other target populations. Additionally, as discussed in Chen and Ho's study, Chinese college students are considered to have a relatively lower difference in language proficiency compared to other groups of the population [9]. Thus, it might account for the unexpected insignificance of several results such as insignificance in the influence of language proficiency, especially considering that our participants are constituted by a large number of students who are graduating from or studying at a university in an English-speaking country. However, for future studies to settle such inconsistency in present studies on the influence of language proficiency on Stroop interference effect, it is suggested to have a group of participants that have unbalanced language proficiency to examine for more significant influence or measure the language proficiency of each participant to factor out the potential influence of the individual differences that might harm the reliability of the results.

Also, the language similarity that is considered to influence the interference effect may be used to explain the inconsistencies between Chen and Ho's study as well as our study and Magiste's study that Chen and Ho's study used two disparate languages- English (alphabetical) and Chinese (orthographical)- while Magiste focused on two similar languages- Swedish and German [9]. This could be a direction for future research to settle the dispute between present studies as well.

The similar Stroop interference effect discovered in pseudo-color-word strings that are misspelled is consequential for future studies that examine the interference with non-color-word strings. In addition, we can do the pseudoword test in Chinese in future studies and then compare the corresponding result with what we have achieved for the English pseudoword test, to see whether the interference effect of pseudoword in English can be extended to other languages such as Chinese, and further evaluate the interference effect of pseudoword when factoring out the proficiency in the

language. We may substitute the minor component of the Chinese character indicating the color, to build a 'single Chinese pseudoword' which looks similar to the original one but does not have the original meaning.

5. CONCLUSION

In conclusion, in each language (i.e., English and Chinese), we found a significant Stroop interference-effect. However, no effect of language proficiency was found, as there was no response time difference between the Chinese Stroop test and English Stroop test in this research. Importantly, we did find the effect of automatic linguistic correcting in Stroop tests, which was supported by the presence of a significant congruency effect and facilitating effect in English pseudowords. The results serve as evidence that people tend to automatically correct non-words and process the meaning of the corresponding real words familiar to them through a top-down process. Supported by this study, this linguistic phenomenon is considered pervasive, influential, highly autonomous, and difficult to inhibit among people's daily cognitive activities.

REFERENCES

- [1] Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18(6), 643.
- [2] Jensen, A. R., & Rohwer Jr, W. D. (1966). The Stroop color-word test: a review. *Acta psychologica*, 25, 36-93.
- [3] Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive psychology*, 12(4), 523-553.
- [4] Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance*, 8(6), 875.
- [5] Scarpina, F., and Tagini, S. (2017). The Stroop color and word test. *Front Psychol*, 8: 557.
- [6] Guo, Q. H. H. Z., Hong, Z., Lv, C., Zhou, Y., Lu, J. C., & Ding, D. (2005). Application of Stroop color-word test on Chinese elderly patients with mild cognitive impairment and mild Alzheimer's dementia. *Chinese Journal of Neuromedicine*, 4(7), 701-704.
- [7] Feng H, Li G, Xu C, et al. (2017). Training rehabilitation as an effective treatment for patients with vascular cognitive impairment with no dementia. *Rehabil Nurs*, 2017; 42: 290-297.
- [8] Wang, R. Y., Zhou, J. H., Huang, Y. C., & Yang, Y. R. (2018). Reliability of the Chinese version of the Trail making test and stroop color and word test among older adults. *International Journal of Gerontology*, 12(4), 336-339.
- [9] Chen, H. C., & Ho, C. (1986). Development of Stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning, memory, and cognition*, 12(3), 397.
- [10] Mägiste, E. (1984). Stroop tasks and dichotic translation: The development of interference patterns in bilinguals. *Journal of Experimental Psychology: Learning, memory, and cognition*, 10(2), 304.
- [11] Erhi, L. C. (2005). Development of sight word reading: Phases and findings. In M. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp.135-154). Oxford, UK: Blackwell Publishing Ltd.