**Stroop effect modulation by concurrent working memory load**

*Introduction*

Cognitive control is a defining characteristic of human behavior and can be described as the ability to act purposefully and flexibly while overriding automatic or immediately appealing behaviors. Thus, cognitive control has a significant role in our abilities to make decisions, plan, and solve problems. In his book on cognitive control Tobias Egner (2017) explains that the concept emerged from the study of communication and control systems, including the mid-20th century field of cybernetics. According to Egner (2012), the seminal book of Miller and colleagues called “Plans and the structure of behavior” (1960) explicitly linked control theory with the goal-oriented nature and hierarchical structure of human cognition.

Understanding the mechanisms that facilitate cognitive control is an important task that could help explain why we are capable of intelligent, purposeful behavior, yet also susceptible to irrational influences and failures (Egner, 2017). The dichotomy between controlled and automatic processing is a fundamental principle in cognitive psychology and despite its importance, the construct of cognitive control has proven elusive, particularly when it comes to understanding the underlying mechanisms (Egner, 2017).

According to Egner (2017), Posner & Snyder (1975) used the Stroop task (Stroop, 1935) as a classic example to demonstrate the difference between controlled and automatic processes. In this task, adults are typically quicker to read a word aloud than to name the color of a stimulus, meeting the first criterion. When faced with incongruent stimuli (like the word 'YELLOW' in gray color), the color of the stimulus weakly affects the word-reading response, but the word invariably disrupts color naming and moreover, if one tries to name the color while engaged in an unrelated task (like mental arithmetic), performance is likely to suffer (Egner, 2017). According to the author, these characteristics do not apply to word reading, which led to the conclusion that color naming is a controlled process, while word reading is automatic. Thus, the interpretation of the results of the Stroop task has become a fundamental framework for studying controlled and automatic processing.

The operationalization of cognitive control highlighted three characteristics that differentiate controlled processes from automatic: first, controlled processes are slower to conduct; second, they can be disrupted by competing automatic processes and third, they depend on a frontal processing mechanism with limited capacity (Egner, 2017).

The involvement of the frontal executive system in managing the Stroop task is further substantiated by studies examining individual differences (Goldfarb & Henik 2007, Kalanthroff et al., 2015, Kalanthroff et al., 2018). A study by Zhao and colleagues (2014), has shown a link between the size of the Stroop effect and a person's working memory capacity, which is another crucial function of the frontal brain regions. Working memory capacity is often measured using the operation span (OSPAN) task (Turner & Engle, 1989) and participants with a lower operation span tend to have longer reaction times in incongruent Stroop trials (Kane & Engle, 2003; Meier & Kane, 2012, as cited in Kalanthroff et al., 2015). This suggests that individuals with a higher working memory capacity are better able to select the relevant information in the Stroop task. However, although these previous studies demonstrated strong support for the connection between a frontal executive system that maintains and updates information in working memory and Stroop interference, they do not establish causality (Kalanthroff et al., 2015).

The aim of this experiment is to partially replicate Kalanthroff and colleagues (2015) experiment and investigate how reducing working memory resources affects performance on a simultaneous Stroop task. Following the design of Kalanthroff and colleagues (2015), I selected the n-back task to engage working memory resources. This task is known for its requirement of both retention and updating of information, as well as it is recognized for activating the frontal executive system (Owen et al., 2005, as cited in Kalanthroff et al., 2015).

In contrast to the original experiment of Kalanthroff and colleagues (2015), employing the n-back task alongside the Stroop task resulted in three levels of difficulty (versus two – 0-back and 2-back): low (0-Back), medium (1-Back) and high (2-Back). Anticipated results were that under the medium and high-load conditions, executive resources would be engaged in maintenance and updating tasks, leaving fewer resources for managing the Stroop task. Reducing working memory resources would primarily impact participants' responses to incongruent Stroop stimuli, as they would require more time to resolve the conflict with fewer top-down resources.

*Method*

*Participants*

Twenty-five psychology students (17 females and 8 males) of New Bulgarian University participated for fulfillment of course requirements and credit. All participants had normal or corrected-to-normal vision and were native speakers of Bulgarian. The mean age was 22.60 years (SD = 3.63). Data from six subjects were removed from the 2-back condition due to an error rate of more than 50%.

*Stimuli and Design*

The experiment employed a within-subjects factorial design with two independent variables: congruency, with two levels (congruent, incongruent), and letter N-Back, with three levels (0-back, 1-back, and 2-back). Three dependent variables were measured: reaction time in milliseconds, N-Back accuracy rate: hit rate (the rate of correct responses to target trials) and correct rejection rate, and Stroop task accuracy: percent correct responses. The stimulus material was composed of one Stroop list containing ten words and three additional lists, each containing sixty letters used in the N-Back tasks. For the Stroop task, each stimulus consisted of one of three Bulgarian color words— (blue), (yellow), and (gray). The colors of the stimuli were blue, yellow, and gray. Three congruent, and six incongruent combinations of colors and words were created with the proportion of 40-60 for congruent and incongruent. For the n-back tasks, a total of 25 Bulgarian consonant letters were used.

*Procedure*

The experiment was conducted in the laboratory room of the department of Cognitive Science and Psychology at New Bulgarian University. It began with practice blocks for each task, with participants performing twenty practice trials per block, starting with the Stroop task. Participants were instructed to respond to the ink color of the word using the keys “1”, “2”, and “3”. Color-key bindings were counterbalanced. Each practice trial began with a fixation cross at the center of the screen for 500-ms, followed by a word representing one of the three colors. The word remained on the screen for 5000-ms or until the participant responded. After each response, a feedback display with accuracy and reaction time (RT) was presented for 1000-ms, followed by a 1000-ms inter-trial interval.

For the n-back practice blocks, participants responded to the letter “Z” (0-back), the letter from the previous trial (1-back), or the letter from two trials earlier (2-back), using the keys “5” and “6” for the same and different. The target letter for all participants during the experimental trials in the 0-back condition was “T”. The appearance of the target and distractor letters in the 0-back condition was randomized, while the target and distractor letters within the 1-and-2 back conditions were presented sequentially. Each n-back practice trial followed the same sequence as the Stroop practice trials.

After the practice trials, a combined Stroop and n-back task started (see the experimental task in Fig. 1) with counterbalanced low (0-back), medium (1-back), and high (2-back) working memory (WM) load conditions. Each participant performed a total of 80 training and 180 experimental trials. Stimuli were presented on a 24-inch, 120Hz refresh rate monitor with 1080p resolution, using E-Prime version 2 (SP2) software. Participants’ responses were recorded through a mechanical keyboard. Stimuli were presented at the center of the screen and appeared in white on a black background in Courier New font with size 24. The viewing distance was approximately 60 cm. The experiment took approximately 20 minutes to complete.

*A picture containing darkness, screenshot, black

Description automatically generatedFig. 1 – Example of experimental trial*

*Results*

Similar to the experiment of Kalanthroff and colleagues (2015), for the n-back task two separate accuracy rates were calculated: hits and correct rejections. For the 0-back condition, the mean hit rate was .962, SD = .06, and the mean correct rejection rate was .995, SD = .01. For the 1-back condition, the mean hit rate was .775, SD = .16, and the mean correct rejection rate was .962, SD = .04. For the 2-back condition, the mean hit rate was .711, SD = .12, and the mean correct rejection rate was .890, SD = .10. The hit rates and correct rejections were in line with the reported accuracies from the experiment of Kalanthroff and colleagues (2015).

One-Way ANOVA demonstrated significant differences within the accuracies of hits (F = 26.87, p < 0.001, η2p = 0.45), and correct rejections (F = 16.30, p < 0.001, η2p = 0.33). Post hoc analysis with Bonferroni adjustment revealed that for hits, the 2-back condition (M = 0.71, p < 0.001) was significantly lower from both 0-back (M = 0.96, p < 0.001) and 1-back (M = 0.77, p < 0.001) conditions. The difference between 0-and-1 back conditions was not significant (p > 0.05). For correct rejections, the 2-back condition was significantly lower from both 0-back (M = 0.99, p < 0.000), and 1-back condition (M = 0.96, p < 0.001). The difference between 0-and-1 back conditions was again not significant (p > 0.05). These results demonstrate that the 2-back condition was more difficult, and hence caused a higher WM load, thus negatively affecting the accuracy of both hits and correct rejections in a significant way.

Across all conditions, the mean overall accuracy of the Stroop task was as follows. In the 0-back condition, accuracy was .97, SD = .19 for congruent and .95, SD = .19 for incongruent. In the 1-back condition the accuracy of congruent was .96, SD = .25, and .90, SD = .25 for incongruent. Within the 2-back condition, the congruent accuracy was .97, SD = .21 and .94, SD = .21 for incongruent. The Stroop mean RTs for correct responses as a function of concurrent WM load (low vs. medium vs. high control) and congruency are presented in Table 1.



*Table 1 - Mean RTs in milliseconds under the different load and congruency conditions of the combined n-back and Stroop tasks.*

Before continuing with additional analyses, RTs from incorrect trials (11.4% of the data) were excluded along with RTs greater than 2.5 SDs above and below the sample mean (an additional 7.8% of the data). The exclusions were applied to both Stroop and n-back conditions.

A 2x3 factorial ANOVA was conducted to examine the effects of both the Stroop congruency (congruent vs. incongruent) and the varying levels of concurrent WM load (0, 1, and 2-back) on Stroop RTs. Results indicated a significant main effect of concurrent WM load on Stroop RTs, F(2, 132) = 4.08, p < 0.05, η²p = 0.06. There was also a significant main effect of Stroop congruency, F(1, 132) = 7.38, p < 0.01, η²p = 0.05, with longer RTs in the incongruent, compared to the congruent condition. The interaction effect between Stroop congruency and the concurrent WM load levels was not significant, F(2, 132) = 0.06, p > 0.05, η²p = 0.00.

Post hoc pairwise comparisons with Bonferroni adjustment were performed on the three levels of the n-back task and Stroop congruency on Stroop RTs. The results indicated significant differences in RTs between 0-back congruent with 1-back incongruent (p < 0.05) and 0-back congruent with 2-back incongruent (p < 0.05). The remaining differences were not significant.

Post hoc pairwise comparisons with Bonferroni adjustment were also performed on Stroop congruency. Results demonstrated that the Stroop effect (Stroop, 1935) was observed, with significant difference in the RTs between the conditions (p < 0.01). Mean RTs were significantly higher (M = 994) for the incongruent condition, compared to the congruent (M = 892) condition.

Correlation analysis was performed to assess the relationships between Stroop congruency, concurrent WM load levels and Stroop RTs. Results revealed significant but weak positive correlation between Stroop congruency, r(138) = .22, p < .01, and concurrent WM load levels, r(138) = 0.23, p < .01 RTs. One-hot encoding the data revealed significant but weak negative correlation between Stroop RTs and the congruent condition, r(138) = -.23, p < 0.01. Significant but weak positive correlation between Stroop RTs and the incongruent condition, r(138) = .23, p < 0.01. Additionally, significant but weak negative correlation was found between Stroop RTs and the 0-back condition, r(138) = -0.21, p < 0.05. Significant but weak positive correlation was found between Stroop RTs and the 2-back condition, r(138) = 0.18, p < 0.05. The correlation between Stroop RT and 1-back condition was not significant.

Multiple regression was conducted to reveal whether Stroop congruency and N-Back levels were significant predictors of Stroop RTs. The model was statistically significant, F(3, 134) = 5.38, p < .005 and explained approximately 10% of the variance in Stroop RTs (adjusted R² = .090). For the N-Back conditions the 0-back was associated with an increase of 104-ms in Stroop RT, b = 104.16, t(134) = 4.14, p < .001. The 1-back was associated with an increase of 179-ms in Stroop RT, b = 179.90, t(134) = 7.156, p < .001. The 2-back was associated with an average increase of 233-ms, b = 233.39, t(134) = 8.458, p < .001. In terms of Stroop congruency, the congruent condition was associated with an increase of 208-ms in RT, b = 208.03, t(134) = 11.099, p < .001, and the incongruent condition was associated with an increase of 309-ms in RT, b = 309.42, t(134) = 16.509, p < .001.

*Discussion*

The results from the experiment supported the hypothesis that reducing WM resources would negatively impact the RTs to incongruent stimuli. Additionally, as the N-Back level increased, so did the Stroop RTs, suggesting an increase in cognitive load. This result is aligned with the notion that cognitive load associated with the N-Back task is likely to detract from the resources available to manage the Stroop task (Kalanthroff et al., 2015). A significant effect of the incongruent condition was observed, which resulted in longer RTs compared to the congruent condition. This aligns with the concept that incongruent stimuli present a conflict that needs resolving, and thus, are more demanding of executive resources (Stroop, 1935). Therefore, under conditions of increased cognitive load due to the N-Back task, the efficiency of conflict resolution in the Stroop task is reduced.

The connection of WM load to Stroop task performance aligns with the understanding that cognitive control is an important element of human cognition and its role in the resolution of interference and conflict. The results from the experiment relate to the theoretical framework of the Dual Mechanisms of Control Theory (DMC; Braver, 2012, as cited in Kalanthroff et al., 2015), which addresses variability in cognitive control during the Stroop task by invoking two types of conflict and control.

According to Braver (2012), Stroop performance relies on two sources of conflict: informational conflict and task conflict. The former emerges from the discrepancy between word meaning and font color, while the latter comes from the competition between the dominant, but contextually irrelevant word-reading task and the less automatic, but contextually relevant color-naming task (Braver, 2012). This suggests that the variations in Stroop task performance under varying WM load can be interpreted as an indication of these two types of conflict. The increase in RTs in incongruent Stroop stimuli under higher WM load might be understood as an increase in informational conflict due to reduced cognitive resources available to resolve the Stroop conflict.

Additionally, the experimental findings relate to the concept of proactive control (PC) in DMC (Braver, 2012). Proactive control refers to the early selection and maintenance of goal-relevant information to guide behavior and is assumed to reduce both types of conflict prior to the occurrence of the Stroop stimuli (Kalanthroff et al., 2018). The theory suggests that when PC is reduced, it results in a distinctive pattern of increased Stroop interference coupled with a reversal of Stroop facilitation, which is also known as the reverse-facilitation (RF) effect (Kalanthroff et al., 2015, Kalanthroff et al., 2018). Although in the current experiment the RF effect was not manipulated and measured, the observed pattern of increased Stroop interference under higher WM load conditions could indirectly indicate a reduction in PC. As WM resources were engaged in the n-back task, fewer resources were available for PC in the Stroop task, potentially resulting in increased informational and task conflict. However, to back this claim the experimental design had to include neutral stimuli in the Stroop task, based on the assumption that the lexical pathway isn't stimulated in a bottom-up manner during neutral trials, which aligns with the utilization of nonword neutral items (see Kalanthroff et al., 2015, Kalanthroff et al., 2018). According to the authors, neutral trials provide a baseline measure of the cognitive resources required to perform the task without the interference (incongruent trials) or facilitation (congruent trials) caused by the relationship between the word and its color.

Overall, the results from the experiment offer empirical support to the DMC theory, suggesting that variations in cognitive load could influence the extent of PC, and thereby modulate the informational and task conflict in the Stroop task.

It is important to consider these results within the context of the experiment’s constraints and the following limitations that should be acknowledged. First, participant sample was relatively small (especially in the 2-back condition), which may limit the generalizability of the findings. Second, there were three levels of the n-back task (0-back, 1-back, and 2-back), while other studies (Kalanthroff et al., 2015) have used only two levels (0-back and 2-back), effectively avoiding the medium WM load condition. These studies also introduced a third (neutral) level in the Stroop task, presented by nonword items (XXXX), which allowed the researchers to measure Stroop reverse-facilitation (RF) and interference (Kalanthroff et al., 2015).

In conclusion, the current experiment provides insights into the effects of n-back task difficulty and stimulus congruency on cognitive processing speed. While evidence of an interaction effect between these factors was not found, these findings may contribute to a growing body of literature on cognitive load and the control of attention.

*References:*

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in cognitive sciences*, *16*(2), 106-113.

Egner, T. (Ed.). (2017). The Wiley handbook of cognitive control. *John Wiley & Sons*.

Goldfarb, L., & Henik, A. (2007). Evidence for task conflict in the Stroop effect. Journal of Experimental Psychology: Human Perception and Performance, 33(5), 1170.

Kalanthroff, E., Avnit, A., Henik, A., Davelaar, E. J., & Usher, M. (2015). Stroop proactive control and task conflict are modulated by concurrent working memory load. *Psychonomic Bulletin & Review*, 22, 869-875.

Kalanthroff, E., Davelaar, E. J., Henik, A., Goldfarb, L., & Usher, M. (2018). Task conflict and proactive control: A computational theory of the Stroop task. *Psychological review*, *125*(1), 59.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of experimental psychology: General*, *132*(1), 47.

Meier, M. E., & Kane, M. J. (2013). Working memory capacity and Stroop interference: global versus local indices of executive control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 748.

Miller, G. A., Galanter, E., & Pribram, K. A. (1960). *Plans and the structure of behavior*. New York: Holt, Rhinehart, & Winston.

Posner, M. I., & Snyder, C. R. R. (1975). In R. L. Solso (Ed.), Information processing and cognition: The Loyola symposium. New York: Lawrence Erlbaum

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18(6), 643.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. Journal of memory and language, 28(2), 127-154.

Owen, A.M., McMillan, K.M., Laird, A.R. and Bullmore, E. (2005), N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies*. Hum. Brain Mapp*., 25: 46-59. https://doi.org/10.1002/hbm.20131

Zhao, Y., Tang, D., Hu, L., Zhang, L., Hitchman, G., Wang, L., & Chen, A. (2014). Concurrent working memory task decreases the Stroop interference effect as indexed by the decreased theta oscillations. *Neuroscience*, 262, 92-106.