The Stroop Competition: A Social-Evaluative Stroop Test for Acute Stress Induction

Victoria Mueller[1*], Robert Richer[1], Lea Henrich[1], Leonie Berger[2], Antonia Gelardi[2], Katharina M. Jaeger[1], Bjoern M. Eskofier[1], and Nicolas Rohleder[2]

***Abstract*—The Stroop test is one of the most widely used protocols to induce cognitive stress and reliably activates the sympathetic nervous system (SNS). However, it only moderately activates the hypothalamic-pituitary-adrenal (HPA) axis, the stress axis responsible for cortisol secretion. In other well-known stress protocols, such as the cold pressor test, adding social-evaluative elements to the regular procedure has proven to cause increased HPA axis activation. For this reason, we introduce the "Stroop Competition", a novel stress protocol based on the established Stroop test that adds social-evaluative feedback by conducting the test against a fake opponent with subsequent performance evaluation. We investigated the stress response of 22 participants performing either the "Stroop Competition" (Competition group) or the regular Stroop test (Control group) three consecutive times. Stress responses were assessed using ECG recordings to extract heart rate (HR) and heart rate variability (HRV) and saliva samples to extract salivary alpha-amylase (sAA) and cortisol. In the Competition group, participants experienced higher SNS activation indicated by significantly higher HR and lower HRV levels as well as higher sAA response to the stressor compared to the Control group. Additionally, overall cortisol levels were significantly higher in the Competition group supporting higher HPA axis activity. The findings of our pilot study confirm our hypothesis that adding social-evaluative elements to the Stroop test causes a more effective activation of both the SNS and HPA axis. We are convinced that our novel "Stroop Competition" protocol will provide a valuable addition to the already existing stress protocols in biopsychological research.**

***Index Terms*—Stroop test, Acute stress, Cortisol, Alpha-amylase, Heart rate variability**

I. Introduction

The Stroop effect is an experimental psychological phenomenon that arises when the processing of a particular stimulus interferes with the simultaneous processing of a second one. This interference effect is used in the *Stroop (Color Word Interference) test*, which was originally proposed by John R. Stroop in 1935 [1]. In such a test individuals either have to read a sequence of color words or name the colors the words are written in. The text color of the word either matches the color name (*congruent*) or does not match (*incongruent*). The incongruent condition creates the challenge to stop the automated task of reading only the written word and instead name correctly the color in which the word is written, causing a cognitive inhibition known as the *Stroop effect*.

*Responsible author; Contact: victoria.m.mueller@fau.de

[1]Machine Learning and Data Analytics Lab (MaD Lab), Department Artificial Intelligence in Biomedical Engineering (AIBE), Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 91052 Erlangen, Germany;

[2]Chair of Health Psychology, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 91052 Erlangen, Germany

In stress-related research, the Stroop test is one of the most widely used protocols to induce cognitive stress [2]. It causes activation of the sympathetic nervous system (SNS) leading to increased heart rate (HR), decreased heart rate variability (HRV), and the secretion of the saliva enzyme alpha-amylase (sAA) [3]. However, the Stroop test induces only moderate increases in cortisol, which is a glucocorticoid secreted by the hypothalamic-pituitary-adrenal (HPA) axis, the second major stress system besides the SNS [4]. Cortisol plays a key role in the human stress reaction and is associated with psychological, physiological, and physical health functioning [5]. Thus, stress protocols that reliably activate the HPA axis are crucial for biopsychological research.

Effective HPA axis activation can be achieved in stress tasks that involve social-evaluative elements, hence inducing psychosocial stress [5]. The gold standard laboratory protocol for acute psychosocial stress induction is the Trier Social Stress Test (TSST, [6]), in which participants have to give a free speech and solve mental arithmetic tasks in front of an evaluative panel while being recorded. However, conducting the TSST is very resource-intensive and limits conducting biopsychological research at a larger scale since it requires at least two trained study leaders serving as evaluation panel and one additional study supervisor carrying out the experiment. For that reason, researchers have attempted to modify established stress protocols to also activate the HPA axis. For instance, Schwabe et al. transformed the cold-pressor test (CPT, [7]), a stress protocol for physiological stress induction, where individuals have to immerse their hand into ice water, into a socially evaluated cold pressor test (SECPT) [8]. In comparison to the CPT, individuals performing the SECPT showed increased HPA axis activity. However, even though the (SE)CPT has been used in a wide range of studies, its general applicability is limited since immersing the dominant hand in cold water can induce pain. Further, it can hinder the collection of other biomarkers for stress assessment that require the hand, such as electrodermal activity (EDA) [9].

Transforming the Stroop test into a social-evaluative stress test might present a better-suited alternative for acute stress induction than the existing protocols. For that reason, we present the "Stroop Competition", a novel stress protocol based on the well-established Stroop test that consists of performing the Stroop test repeatedly against a privy opponent with subsequent performance evaluation. We hypothesize that our adaptation of the Stroop test leads to a stronger activation of both the SNS and HPA axis. To validate our approach we assess electrophysiological and salivary biomarkers during the

Stroop Competition and a regular, non-competitive Stroop test and compare them. To the best of our knowledge, this is the first version of a social-evaluative Stroop test.

II. Methods

A. Data Acquisition

To evaluate the Stroop Competition we collected data from $n = 22$ healthy participants (50 % female) aged 24.4 ± 7.2 years ($\mathrm{M} \pm \mathrm{SD}$) years during the COVID-19 pandemic in December 2020. Participants were asked to get up at least three hours prior to the study to minimize the impact of circadian variations in hormone concentrations [10] and to not perform any vigorous physical activity at least 3 hours before the study. Additionally, they were instructed to avoid the consumption of alcohol on the day of the study, and of food and beverages (except water and unsweetened tea) at least 1 hour before the beginning of the study. The participants were randomly assigned to the *Competition* ($n = 10$) or the *Control* group ($n = 12$). All participants gave written informed consent before the study. The study protocol was approved by the Ethics Committee of Friedrich-Alexander-Universität Erlangen-Nürnberg (number 106_13B).

The stress test consisted of three repetitions of the Stroop test on a computer. The test was implemented using Inquisit 6 (Millisecond Software, Seattle, WA, USA) and consisted of 40 color words that were randomly chosen to be either congruent or incongruent. Participants used the keyboard to respond to the displayed color words which were either red, green, blue, or black. Auditory and visual feedback was provided in form of a green checkmark for correct answers and a red cross for incorrect answers, accompanied by a representative tone. For the Competition group, participants were informed that they had to perform the Stroop test against an opponent directly after providing informed consent. They did not know that the opponent was privy and part of the study supervisors. To ensure a constantly better Stroop performance the opponent only had to perform a purely congruent Stroop test with 36 words. To elicit additional social-evaluative pressure we provided feedback to the Competition group after each Stroop test in form of a fake statistic, where their performances were constantly worse than the opponents'.

Throughout the procedure, we recorded ECG data and collected saliva samples to assess SNS and HPA axis activation. ECG data were recorded using a wearable ECG sensor node (Portabiles GmbH, Erlangen, Germany), attached to a chest strap, recording a 1-channel ECG according to Lead I of Einthoven's Triangle with a sampling frequency of 256 Hz onto the internal storage for subsequent data processing on a computer. Before the Stroop test, we recorded 10 min of ECG baseline. The three repetitions of the Stroop test constitute the three phases used for ECG data analysis: *Stroop1*, *Stroop2*, and *Stroop3*. Cortisol and sAA were assessed by collecting four saliva samples using Salivettes (Sarstedt AG & Co. KG, Nümbrecht, Germany). Participants were asked to chew on a polystyrol roll for one minute. The first saliva sample (*S0*) was taken directly on arrival ($t = -15$ min relative to Stroop test start), followed by three samples (*S1-S3*) $t = \{0, 10, 20\}$ min relative to the start of the Stroop test. After collection, saliva samples were stored at −18 °C for later analysis.

B. Data Processing

1) ECG: The acquired ECG data were used to extract RR intervals after filtering and applying a QRS detection algorithm provided by the *Neurokit2* library [11]. Artifacts in RR intervals were reduced according to previous work (e.g., [12]). From the extracted RR intervals, we computed heart rate (HR) and the HR increase relative to Baseline in percent (ΔHR) to assess participants' individual physiological reactions. Additionally, we computed the HRV measures RMSSD, pNN50, and SD1/SD2 according to HRV taskforce recommendations to characterize sympathetic activation [13].

2) Biomarker: Salivary cortisol concentrations were determined in duplicate using a chemiluminescence immunoassay (CLIA, IBL, Hamburg, Germany) as described in previous publications (e.g., [14], [15]) after centrifuging the collected saliva samples at 2000 g and 20 °C for five minutes. sAA was measured with an in-house enzyme kinetic assay using reagents from Roche Diagnostics (Mannheim, Germany), as previously described [16]. Besides raw cortisol and sAA values, we computed the areas under the curves (AUC_{cort} and AUC_{sAA}) as measures for the total amount of cortisol and sAA secretion, respectively. Missing *S1* or *S2* samples ($n = 3$ participants) were imputed by linear interpolation.

III. Statistical Analyses

For statistical analyses, we first tested the data for normal distribution (Shapiro-Wilk test) and for homogeneity of variances (Levene test). We then performed mixed-measurement analyses of variance (Mixed-ANOVA) to test for interaction effects between condition (between-factor) and Stroop phases (within-factor). As post-hoc tests, we used pairwise t-tests with Bonferroni correction for multiple-comparison correction. For the AUC_{cort} and AUC_{sAA} we performed t-tests between the two groups. All statistical analyses were performed at a significance level of $\alpha = 0.05$. Effect sizes of ANOVA are reported as η_p^2 and of t-tests as Cohen's d. To indicate statistical significance in Figures and Tables we used the the following notation: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. All analysis steps were performed using *BioPsyKit* [17], an open-source package for the analysis of biopsychological data.

IV. Results

A. ECG

On average, HR throughout all three Stroop phases was 32.1 % higher for the Competition condition. This was supported by a statistically significant interaction effect between Stroop phase and group, $F(2, 36) = 4.132, p = 0.024, \eta_p^2 = 0.187$. Post-hoc testing revealed significant differences between both groups in all Stroop phases.

Similarly, the relative HR increase compared to *Baseline* (ΔHR) was higher for the Competition than for the Control group throughout all Stroop phases (Figure 1). Mixed-ANOVA

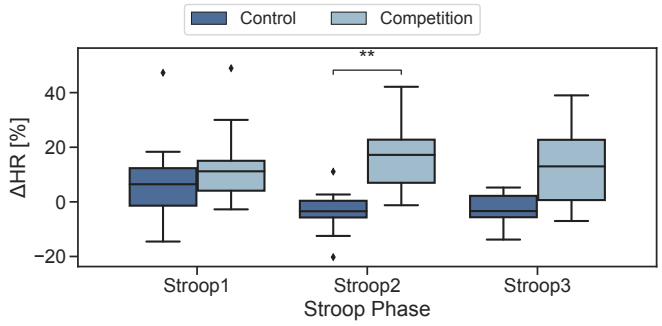

Fig. 1. HR increase relative to *Baseline* per *Stroop phase* and group.

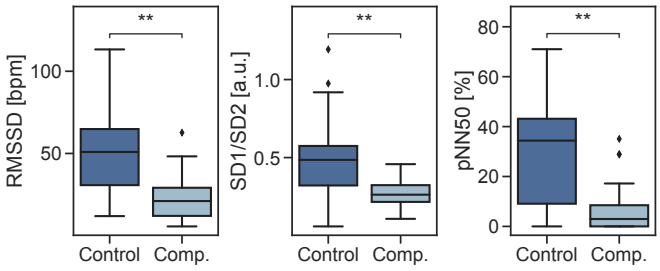

Fig. 2. HRV measures per group during Stroop test.

revealed a significant interaction between group and phase, $F(2, 36) = 3.726, p = 0.034, \eta_p^2 = 0.172$. Post-hoc testing showed that ΔHR was significantly higher for the Competition group during *Stroop2* (Figure 1). For the HRV measures, mixed-ANOVA did not show any interaction effects. However, the main effect *group* was significant for each HRV measure, indicating lower HRV for the Competition group (Figure 2).

B. Saliva

Results show that initial sAA levels (*S0*) were higher for the Competition group and reached similar levels right before the beginning of the Stroop test (*S1*) (Figure 3, left). However, the Competition group had a considerably higher sAA increase in response to the Stroop test compared to the Control group.

Similarly, the Competition group had overall higher cortisol levels across all time points (*S1*) (Figure 3, right). This is supported by a significant group main effect, $F(1, 18) = 5.556, p = 0.03, \eta_p^2 = 0.236$ and by a significantly higher AUC_{cort} $(t(14.078) = 2.362, p = 0.033, d = 1.099)$. In comparison to the Control group, the Competition group showed higher cortisol increases between *S0* and *S1* as well as between *S2* and *S3*.

V. Discussion

The main objective of our study was to investigate whether adding social-evaluative elements to the Stroop test can lead to a higher stress response resulting in increased SNS, and, especially, increased HPA axis activity. Our results confirm previous findings that the regular Stroop test as well as the Stroop Competition both reliably activate the SNS and thus reproduce results from previous work (e.g., by Hjemdahl et

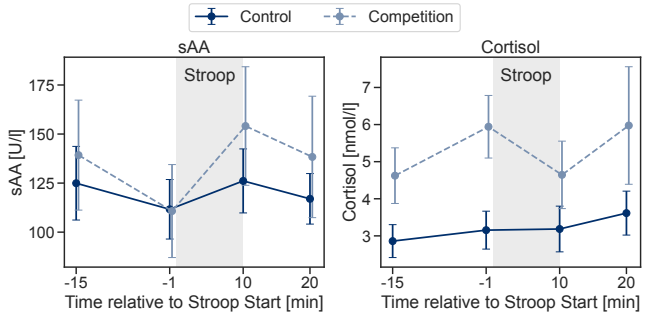

Fig. 3. Raw sAA (*left*) and cortisol (*right*) values for each group. Values are depicted as mean ± standard error.

al. [18]). Participants in the Competition group, however, showed significantly higher HR levels during all three Stroop phases. This is further supported by a stronger HR response to the Stroop test (Figure 1), indicated by ΔHR, and by significantly lower HRV measures compared to the Control group (Figure 2). Our findings confirm our hypothesis that the social-evaluative setting, by competing against an opponent, leads to higher SNS activation. Apparently, participants were influenced by the performance of the privy opponent who always answered faster and made fewer mistakes. Moreover, the false statistics displayed after each trial further affected the subject, as this enhanced the impression that the opponent had performed much better. In contrast, the Control group was able to adapt to the difficulty of the Stroop test, resulting in a constant HR decrease. Similar findings can also be observed in sAA. While initial sAA levels were comparable for both conditions, the sAA increase from *S1* to *S2* was considerably higher in the Competition group. This further supports higher SNS activation due to the competitive situation.

In addition to increased SNS activation, our results also indicate higher HPA axis activity when a social-evaluative situation was created. Overall, cortisol levels were significantly higher for the Competition group than for the Control group. Since the HPA axis reacts slower to a stressing stimulus than the SNS, cortisol levels started to increase 10 min after end of the Stroop test (*S3*) in comparison to sAA which increased immediately after the end of the Stroop test (*S2*) and started to recover again afterwards (Figure 3).

While these results are very promising, there are, however, also some limitations to our study which need to be addressed in future work. Overall, we observed a high standard deviation in cortisol and sAA levels. Due to the limited sample size, we did not exclude subjects with missing saliva samples but imputed missing values by linear interpolation. However, using this approach, we might have missed cortisol or sAA peaks.

Furthermore, the participants in the Competition group were introduced to the opponent immediately upon arrival. This did not allow us to collect a ”true” baseline without any stressing elements and might already have introduced a bias to the general stress level of participants. This effect can, presumably, be observed in the cortisol samples before the Stroop test.

Cortisol levels at *S0*, which were collected after having been introduced to the study protocol and after having provided informed consent (approx. 10 min after arrival) were higher for the Competition group. Cortisol levels then further increased between *S0* and *S1*, whereas no increase was found in the Control group. Additionally, the HR of participants in the Competition group was considerably higher before and during the Stroop test. However, the Stroop Competition still caused higher HR responses which are reflected by significantly higher ΔHR during the Stroop test for the Competition group. To overcome this limitation and to allow the assessment of a true baseline, we plan to adapt our study protocol and inform the participants about the competitive situation only immediately before the beginning of the Stroop test.

Besides, male participants in the Control group may also have been exposed to social evaluation up to some extent since all experimenters were female [19]. However, this potential effect is considerably lower than the social evaluation in the Competition group. Additionally, participants were randomly assigned to each condition, leading to balanced gender distribution in both groups. Another limitation of the study was that the Stroop Competition might not have posed personal relevance for the participants, as we did not announce any prize for the winner of the competition. To generate greater motivation in future work, it might be important to create the prospect of added value for participants. Moreover, participants were unable to win the overall competition after losing the first two Stroop tests to the opponent. Hence, they showed less motivation in the last phase which is reflected by lower ΔHR compared to the second Stroop phase. To generate more motivation, and concurrently induce more stress, the study design can be adapted in order to let the participants win the second round, making the last round the all decisive round.

VI. Conclusion and Outlook

In this work, we presented the "Stroop Competition", a novel stress protocol for inducing acute psychosocial stress. By transforming the established Stroop test into a competitive and evaluative challenge against a privy opponent, we were able to achieve a better activation of both the SNS and HPA axis, the two major stress systems of the human body. Our results show that the Competition group had significantly higher HR responses and considerably higher sAA responses, overall indicating increased SNS activation. The significantly higher cortisol levels in the Competition group indicate increased HPA axis activation. Even though the sAA and cortisol responses to the Stroop Competition did not reach statistical significance between the two groups we observed clear indications that support the general efficacy of our approach. With our pilot study, we were able to identify points of improvement which we will address in future work. Utilizing our findings, we will adapt our study protocol and perform the "Stroop Competition" with a larger number of participants in order to fully validate our approach. We are confident that our novel stress protocol will be an easy-to-implement stress protocol to reliably activate both SNS and HPA axis. Ultimately, we hope that the "Stroop Competition" will provide a valuable extension to the set of already existing stress protocols in biopsychological research.

Acknowledgments

This work was partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1483 – Project-ID 442419336, EmpkinS. Further, B.M.E. gratefully acknowledges the support of the German Research Foundation (DFG) within the framework of the Heisenberg professorship programme (grant number ES 434/8-1).

References

[1] J. R. Stroop, "Studies of interference in serial verbal reactions," *Journal of Experimental Psychology*, vol. 18, no. 6, pp. 643–662, 1935.

[2] A. Alberdi *et al.*, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," *Journal of Biomedical Informatics*, vol. 59, pp. 49–75, Feb. 2016.

[3] R. T. Chatterton *et al.*, "Salivary α-amylase as a measure of endogenous adrenergic activity," *Clin Physiol*, vol. 16, no. 4, pp. 433–448, Jul. 1996.

[4] J. Tulen *et al.*, "Characterization of stress reactions to the Stroop Color Word Test," *Pharmacology Biochemistry and Behavior*, vol. 32, no. 1, pp. 9–15, Jan. 1989.

[5] S. S. Dickerson and M. E. Kemeny, "Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research," *Psychol. Bull.*, vol. 130, no. 3, pp. 355–391, 2004.

[6] C. Kirschbaum *et al.*, "The 'Trier Social Stress Test' – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting," in *Neuropsychobiology*, vol. 28, 1993, pp. 76–81.

[7] E. A. Hines and G. E. Brown, "A standard stimulus for measuring vasomotor reactions: Its application in the study of hypertension," *Proc. Staf Meet Mayo Clin.*, vol. 7, 1932.

[8] L. Schwabe *et al.*, "HPA axis activation by a socially evaluated cold-pressor test," *Psychoneuroendocrinology*, vol. 33, no. 6, pp. 890–895, Jul. 2008.

[9] S. Gradl *et al.*, "An Overview of the Feasibility of Permanent, Real-Time, Unobtrusive Stress Measurement with Current Wearables," in *Proc of the 13th EAI Int Conf Pervasive Comput Technol for Healthc - PervasiveHealth'19*. ACM Press, 2019, pp. 360–365.

[10] J. M. Smyth *et al.*, "Individual differences in the diurnal cycle of cortisol," *Psychoneuroendocrinology*, vol. 22, no. 2, pp. 89–105, 1997.

[11] D. Makowski *et al.*, "NeuroKit2: A Python toolbox for neurophysiological signal processing," *Behav Res*, vol. 53, no. 4, pp. 1689–1696, Aug. 2021.

[12] J. Happold *et al.*, "Evaluation of Orthostatic Reactions in Real-World Environments Using Wearable Sensors," in *2021 43rd Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 6987–6990.

[13] Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J*, vol. 17, pp. 1043–1065, 1996.

[14] J. Janson and N. Rohleder, "Distraction coping predicts better cortisol recovery after acute psychosocial stress," *Biol. Psychol.*, vol. 128, pp. 117–124, 2017.

[15] L. Abel *et al.*, "Classification of Acute Stress-Induced Response Patterns," in *Proc of the 13th EAI Int Conf Pervasive Comput Technol for Healthc - PervasiveHealth'19*. ACM Press, 2019, pp. 366–370.

[16] U. M. Nater and N. Rohleder, "Salivary alpha-amylase as a non-invasive biomarker for the sympathetic nervous system: Current state of research," *Psychoneuroendocrinology*, vol. 34, no. 4, pp. 486–496, 2009.

[17] R. Richer *et al.*, "BioPsyKit: A Python package for the analysis of biopsychological data," *JOSS*, vol. 6, no. 66, p. 3702, Oct. 2021.

[18] P. Hjemdahl *et al.*, "Differentiated sympathetic activation during mental stress evoked by the Stroop test," *Acta Physiol Scand Suppl*, vol. 527, pp. 25–29, 1984.

[19] M. R. Leary *et al.*, "Self-presentation can be hazardous to your health: Impression management and health risk." *Health Psychology*, vol. 13, no. 6, pp. 461–470, Nov. 1994.