



NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
DEPARTMENT OF PHILOSOPHY AND HISTORY OF SCIENCE
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS
DEPARTMENT OF NURSING
DEPARTMENT OF PHILOSOPHY, PEDAGOGY AND PSYCHOLOGY
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS
DEPARTMENT OF INFORMATICS

“Study of Intratask Components of the Stroop Task:
Column time and pause time”

by

Laoura Ziaka

Student Registration Number: 10M05

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science in “Basic and
Applied Cognitive Science” at the Department of
Philosophy and History of Science

Supervisory Committee:

A. Protopapas

N. Smyrnis

Athens

September, 2014

Examination Sheet

A. Protopapas Associate Professor University of Athens

N. Smyrnis Assistant Professor University of Athens

The author asserts that the content of this project is the result of personal work that has been conducted, and the appropriate reference to the work of others, whenever necessary, has been done, in accordance with the rules of academic ethics.

Acknowledgements

I would like to thank my thesis supervisor and Director of the Graduate Program in “Basic and Applied Cognitive Science”, Athanassios Protopapas, for his guidance and support the last three years. Conducting and discussing experiments with him, was an amazing experience and a real pleasure. Special thanks to Professor, N. Smyrnis, for becoming member of my supervisory committee and helping me to approach some of my original research goals.

Many thanks to research fellow, Argyro Vataki, to fellow student of the graduate program, Theodora Dimopoulou, and doctoral candidates, Fotis Fotiadis, Angeliki Altani, and Sofia Loui for their constant support, their constructive comments, and assistance in the present study. Last but not least, I would like to thank all of the participants—adults and children—, without whom these studies would not have been possible.

This thesis is especially dedicated to my parents, Nikos and Vasiliki; my sister, Mairi; my nieces, Myrsini and Lida; and my dear friend, Vangelis. Words lose their meaning, when referring to their support and patience.

Abstract

In two studies, we aimed to investigate intratask components related to the Stroop task. In Study 1, the performance in each column of the card version of the task was examined, and compared to the single-item version, in order to test two conjunctive theories. According to the conflict monitoring hypothesis, performance improvements should be observed, due to control adjustments. In contrast, the strength model of self-control suggests that participants get depleted, as the task progresses, and, consequently, performance impairment should be noticed. Our results showed a steep increase in color naming time during the card version of the task, supporting the latter theory, and contradicting the former. The implications of these findings for Stroop interference and modeling of it are being discussed. In Study 2, we tested the predictions of two proposals regarding pause time in rapid naming tasks, that is, the silent gap between articulations, comparing two single-item versions of the Stroop task, and examining their relationship to reading and naming. Our findings support partially the proposal for retrieval reflected in pause times, and contradict the suggestion for pause time being an inhibition manifestation. Identification and selection of the correct response, as alternatives for pause-time interpretation, are being discussed.

Keywords: Stroop Interference, cognitive control, conflict monitoring, ego depletion, pause time

Περίληψη

Σε δύο έρευνες, θελήσαμε να ερευνήσουμε συνιστώσες που σχετίζονται με το έργο Stroop. Στην Έρευνα 1, εξετάσαμε την επίδοση σε κάθε στήλη στην εκδοχή της καρτέλας του έργου, και τη συγκρίναμε με την μεμονωμένη εκδοχή, προκειμένου να ελέγξουμε δύο εναλλακτικές θεωρίες. Σύμφωνα με την υπόθεση παρακολούθησης της σύγκρουσης, θα πρέπει να παρατηρηθεί βελτίωσης της επίδοσης, εξαιτίας προσαρμογής του ελέγχου. Αντίθετα, το μοντέλο της δύναμης αυτό-ελέγχου προτείνει ότι οι συμμετέχοντες εξαντλούνται, καθώς το

έργο προοδεύει, και, συνεπώς, μείωση της επίδοσης θα πρέπει να παρατηρηθεί. Τα αποτελέσματα έδειξαν μία απότομη αύξηση του χρόνου κατονομασίας κατά τη διάρκεια της καρτέλας υποστηρίζοντας τη δεύτερη θεωρία, και αντικρούοντας την πρώτη. Οι συνέπειες των ευρημάτων μας για την παρεμβολή Stroop και τη μοντελοποίηση της συζητούνται. Στην Έρευνα 2, ελέγξαμε τις προβλέψεις δύο προτάσεων που αφορούν στον χρόνο παύσης, δηλαδή το σιωπηρό κενό μεταξύ αρθρώσεων σε έργα ταχείας κατονομασίας, συγκρίνοντας δύο μεμονωμένες εκδοχές του έργου Stroop, και εξετάζοντας τη σχέση τους με την ανάγνωση και την κατονομασία. Τα ευρήματά μας υποστηρίζουν μερικώς την πρόταση ότι ο χρόνος παύσης αντανακλά ανάσυρση, και αντικρούει αυτή που αντιμετωπίζει τον χρόνο παύσης ως εκδήλωση αναστολής. Η ταυτοποίηση και επιλογή της σωστής απόκρισης, ως εναλλακτικές για την ερμηνεία του χρόνου παύσης, συζητούνται.

Λέξεις-κλειδιά: Παρεμβολή Stroop, γνωστικός έλεγχος, παρακολούθηση της σύγκρουσης, εξάντληση του εγώ, χρόνος παύσης

Table of Contents

Stroop Interference: An introduction	8
Study 1: Column- Time performance	9
Cognitive Control and Conflict Monitoring: <i>Botvinick's model</i>	9
Ego- Depletion, Self-Control Model, and Stroop	14
Card versus Single-item Version	20
Evidence for Adjacent Stimulus Interference in Both Conditions	24
Rationale of the Present Study	29
Experiment 1	31
Method	31
Results	33
Discussion	37
Experiments 2A and 2B	38
Experiment 2A	40
Method	40
Results	43
Experiment 2B	56
Method	56
Results	56
Discussion	68
General Discussion	69
Study 2: Pause Time	77
RAN-Color and Stoop task: Twins, siblings, or distant relatives?	77
Aim of the Present Study	85
Experiment	86

Method **86**

Results **91**

Discussion **98**


Appendix: Appendix A, Appendix B **106**

References **116**

Stroop interference: An introduction

The Stroop task is one of the best known and most studied tasks in the domain of experimental and cognitive psychology. Although its simplicity is quite remarkable, its contribution to understanding concepts like attention, automaticity, executive functions, learning, memory, and conflict monitoring is of major importance.

Its origin is bound with the rise of experimental psychology since a first version of the task was studied in 1886 by Cattell under the supervision of Wilhelm Wundt. Specifically, Cattell measured naming response times to a variety of stimuli and noticed that it took longer to identify objects and features of objects (e.g., color) than to read the corresponding words of these objects and features. He explained his finding in regard to automaticity claiming that due to extensive practice in reading the printed word is immediately associated with its name while in the case of pictures and colors, which are less practiced dimensions, more effort is needed to choose the corresponding name (MacLeod, 1991, 2000).

Fifty years later, John Ridley Stroop (1935) introduced an equivalent to Cattell's task, now known as the Stroop task. In the second experiment of his dissertation, Stroop combined words and colors in such a way that the two dimensions were incompatible. An example of the task is the word "green" printed in red ink (**green**). Participants' task was to name the color of the stimuli as quickly as possible, and to correct errors. This experimental condition was compared with a control condition that consisted of colored rectangles (e.g., ). The stimuli of both conditions appeared in two separate cards of 100 stimuli (10 rows and 10 columns), five colors were used (red, blue, green, brown, purple) and participants were instructed to name the colors from left to right. Stroop measured the total naming time for each card with a stopwatch. What he observed was that it took much longer to respond to the

incongruent stimuli than to the control stimuli. The difference in response times between the two conditions (incongruent minus control) now bears his name, namely Stroop interference. Stroop adopted Cattell's practice account and the notion of automaticity to explain the slowing in naming times in the incongruent condition.

Since the first introduction of the Stroop task, many variants have been developed, for example the counting Stroop task, the emotional Stroop task, and the picture word task. More importantly, an alternative to the original card version of the color word Stroop task has been introduced, namely the single-item version of the task, which is now broadly used because it is thought to have many advantages over the multiple-item card version (response time for each item, error exclusion; for more details see section "Card-version vs. single-trial version").

Currently, both test formats are used as a tool for studying a variety of cognitive and not only mechanisms, among them cognitive control and ego depletion.

Study 1: Column- Time Performance

Cognitive Control and Conflict Monitoring: *Botvinick's Model*

Cognitive control is implemented in challenging and typically conflicting tasks in order to adapt to them and perform them successfully. One critical question for the literature of cognitive control is how the cognitive system is able to determine how much control is needed in order for a task to be accomplished. Detection of situations that require cognitive control is a central aspect of human behavior (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Botvinick, Cohen, & Carter, 2004).

Botvinick et al. (2001) proposed the *conflict monitoring hypothesis* to answer the aforementioned question. According to the authors, a function exists, namely conflict monitoring, which is responsible for detecting and evaluating a potential conflicting situation. After conflict is confirmed, this system "informs" the control centers and

gives rise to behavioral adjustments and adequate information processing to prevent performance decrements.

The idea of conflict monitoring is supported by neuroscientific evidence focusing on a brain region called Anterior Cingulate Cortex (ACC), the frontal part of the cingulate cortex surrounding the frontal part of the corpus callosum. Botvinick et al. (2001) reviewed evidence that enhanced the assumption of ACC's involvement in the proposed monitoring system. For example, Carter et al. (1998) examined the activation of ACC in variants of the Continuous Performance Test¹ designed to produce increased errors and response competition using functional magnetic resonance imaging (fMRI). What they found was that ACC activity was present not only in errors, but also in *correct* responses, suggesting an ACC activity in the presence of error prone situations and implying that ACC contributes to conflict monitoring. Similar results were also found using a variant of the Eriksen flanker task where the target was always flanked by distractors that were either incompatible (e.g., >><<>>) or compatible (e.g., <<<<<<) with the central target. ACC activity was observed during trials of high conflict, indicating an enhanced role of ACC in conflict monitoring (Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999).

The Stroop task has been considered an ideal candidate for the investigation of conflict monitoring and ACC activation, since, due to the nature of the task, conflict monitoring is required to override the automatic reading response and consequently may probe ACC. A lot of neuroscientific work has been conducted examining ACC activity during the incongruent condition. Pardo, Pardo, Janer, and Raichle (1990) were the first to observe increased ACC activation during the incongruent condition of the Stroop task using Positron Emission Tomography (PET). Since then others have confirmed ACC activity during the Stroop task. Carter, Mintun, and Cohen

¹A task where participants are asked to maintain their focus over a period of time in order to respond to a target, although the task is non-challenging and repetitive.

(1995), using PET, also observed ACC activation during the incongruent condition of the single-item Stroop task. More interestingly, Egner and Hirshi (2005) found that successive incongruent trials (iI) are less influenced by distracting words than incongruent trials that are preceded by a congruent one (cI) (see Figure 1.1). In addition, in cI trials ACC activation was enhanced when compared to iI trials. This finding, although more recent, was taken into account by Botvinick et al. (2001) when proposing the conflict monitoring hypothesis, because this tendency was evident in other conflicting tasks such as the Eriksen Flanker Task and also because a frequency effect² had been observed in the Stroop task.

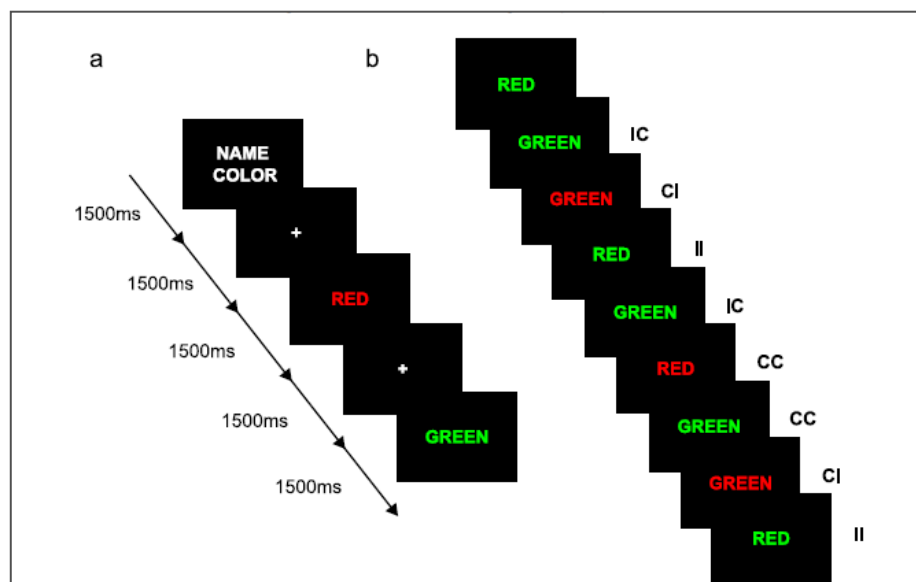


Figure 1.1. Example of the task of Egner & Hirschi (2005) for examining the Gratton effect³ in the Stroop task.

More specifically, and in light of the abovementioned findings, Botvinick et al. (2001) reported two computational models in order to support the conflict monitoring hypothesis and explain neuroscientific and behavioral data. Three of the main findings that are of particular interest for the present study are: a) interference in the Stroop task seems to be greater on the initial trials than on subsequent trails (Henik, Bibi,

²Frequency effect: Reaction times tend to be shorter when incongruent trials are frequent than when they are rare.

³Gratton effect: Reaction times tend to be shorter if incompatible trials are sequential.

Yanai, & Tzelgov, 1997), b) less interference is observed when incongruent trials are frequent than when they are infrequent, and c) sequential adjustments of control seem to be present in the Eriksen flanker task. It should be noted and kept in mind that, due to convenience and practical issues, studies examining frequency effects, performance improvements, control adjustments, and ACC activation during the Stroop task adopt the single-item version of the task, not the card version.

To simulate the Stroop task, Botvinick et al. (2001) adapted the model of Cohen and Huston (1994) and added a conflict monitoring unit, which primarily corresponds to the role of ACC (Figure 1.2). The model has two input layers, one for the color of the stimulus and one for the word, which are connected to an output layer for potential responses. The task demand (or control) units correspond to color naming and word reading, and their activation is responsible for the response. The conflict monitoring unit is connected to the rest of the network and receives input from it. Conflict is defined as “the simultaneous activation of mutually inhibiting units” (p. 630). When incompatible units (i.e., color task demand units vs. word demand units) are inactive the energy of the model is equal to zero and no conflict occurs. The same is true in the case of only one active unit. In contrast, when both incompatible units are active energy increases, indicating conflict. The level of conflict depends on the level of activation of both units: if maximal, the conflict is strong. There is also a feedback loop (connection) between the conflict monitoring unit and the task demand units. As a result of the model’s architecture, in frequent incongruent trials continuous activation of the conflict monitoring unit raises the activation of the control units (i.e., strengthens control and improves performance), while the opposite is true for infrequent incongruent trials (activation of the conflict monitoring unit declines → activation of the control units drops → lax control).

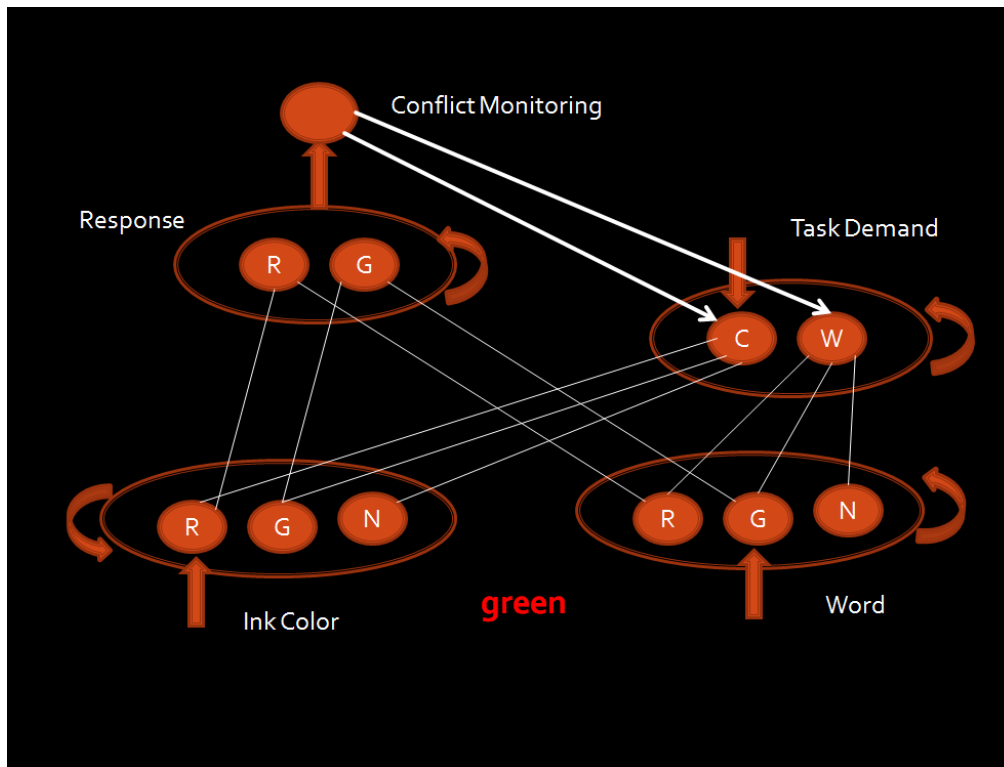


Figure 1.2. Simulation of the Stroop task by Botvinick et al. (2001)

An analogous model was constructed for the Eriksen Flanker Task, where it is supposed that interference is caused due to the nearby stimuli (Figure 1.3). The only difference between the two models is that input units were assigned to the spatial location of the stimuli and the task demand units were replaced by attentional units. Botvinick et al. (2001) managed to successfully simulate all the critical findings regarding the Stroop and the Eriksen Flanker Task, that is, the sequential adjustments in the Eriksen Flanker Task and the trial type frequency effects and improvement performance in the single-item Stroop task. Consequently, this work supported the notion of conflict monitoring and the contribution of ACC to it.

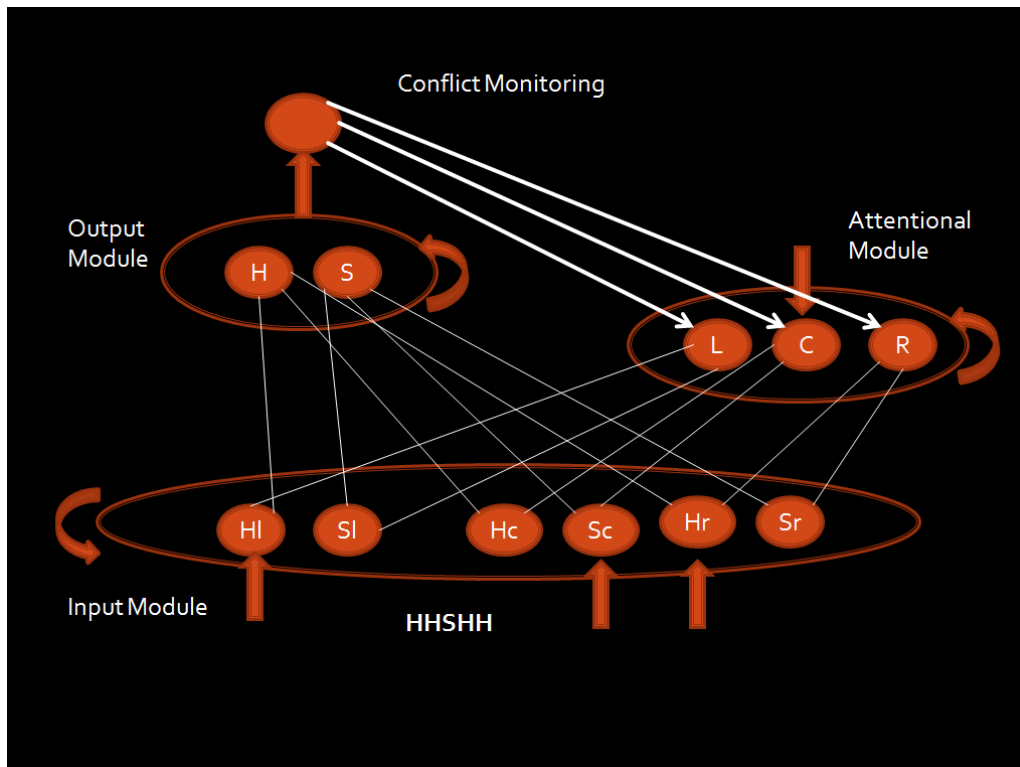


Figure 1.3. Simulation of the Eriksen flanker task.

The above model remains influential and inspiring in the cognitive control literature and is supported from a variety of findings (for review and discussion see Botvinick, Carter, & Cohen, 2004; Carter & van Veen, 2007). In light of the aforementioned findings and the assumption that Stroop performance improves as the task progresses, it is a striking observation that the same task is used in a different literature for the opposite purpose: namely, in the ego-depletion literature the Stroop task is used to induce fatigue and depletion to the participants, leading to gradual performance decline.

Ego Depletion, Self-Control Model, and Stroop

The strength model of self-control has its origin in social psychology, as has its conceiver, namely Roy Baumeister. The central goal of this model is to approach instances where the capability to refrain from an impulsive behavior fails. More specifically, it is thought that goal-oriented and socially acceptable behavior requires from human agents to regulate themselves and inhibit immediate urges or desires. Self-control refers to the capacity to control the self and to the attempt to alter a

thought, a feeling, or a behavior, if required. Following rules and overriding impulses demands from the individual to exert self-control (Muraven & Baumeister, 2000). Although inhibiting a desired behavior is in principle effective, there are instances of failure, indicated, for example, by obesity, drug abuse, eating disorders, and others. Baumeister and his colleagues proposed the strength model of self-control to explain such behaviors. According to this model, when a person is involved in a situation that requires self-control in two consecutive acts, the second act of self-control is impaired. In this model self-control resembles a muscle whose energy is consumable. After the self-control resource has been consumed, a person enters a state called *ego depletion*, where the exertion of self-control is impaired (Baumeister, Vohs, & Tice, 2007; Figure 1.4).

There are several key assumptions deriving from the model:

- 1) Self-control strength is crucial for the function of an individual, that is, for decision-making and for refraining from a behavior.
- 2) A person can only override a limited number of desires at the same time. Self-control strength has a limit, which cannot be extended.
- 3) Each person may have a different level of self-control strength and different tasks demand different levels of self-control.
- 4) Different acts of self-control draw from a common resource. When a person is depleted by a task and performance decrements are observed, these decrements will persist in a second task, even a different one.
- 5) Depletion is not a permanent state. Self-control strength needs replenishment through rest. If rest is prevented, depletion may persist. In other words, continuous efforts of self-control gradually deplete the self-control strength, leading to a performance decrement not only in the current task but also in a subsequent one.

(Baumeister, 2002; Baumeister, Vohs, & Tice, 2007; Muraven & Baumeister, 2000).

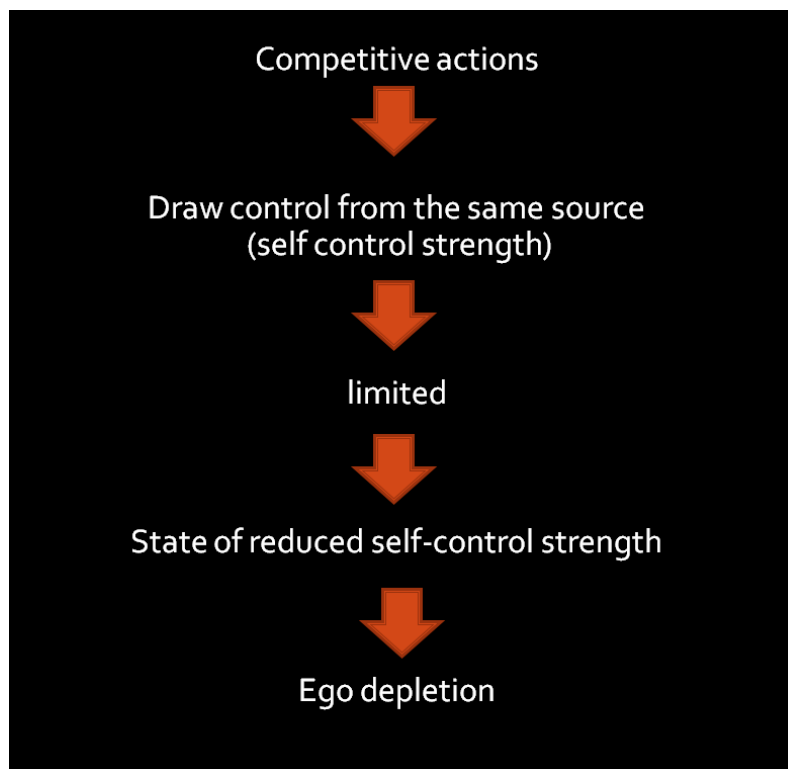


Figure 1.4. Summary of the strength model of self-control.

A standard experimental procedure to test this phenomenon is the dual-task paradigm, in which participants are engaged in two consecutive, unrelated self-control tasks. Participants in the ego depletion group are required to exert self-control in two tasks in a row, both of which demand self-control (typically, the first phase is called depletion phase and the second one manipulation check phase). On the other hand, participants in the control group need to exert self-control only in the second task; their first task is a simpler variant of the first task of the ego-depletion group (again, a control phase is followed by a manipulation check phase). According to the strength model of self-control, participants assigned to the ego depletion group will exhibit impaired performance in the second task, compared to the control group, and that is because the two acts of self-control draw from a global limited resource, which can be depleted (Baumeister et al., 2007). Depleting tasks frequently adopted in the dual-task paradigm are those that require active and effortful inhibition of a dominant and

habitual response or the suppression of an impulse (Hagger, Wood, Stiff, & Chatzisarantis, 2010).

Baumeister (2002) proposed that self-control failures indicate conservation rather than exhaustion, meaning that self-control cannot be fully depleted and people tend to conserve their energy for future self-control acts. If people expect to get involved in another act of self-control soon after the first self-control act, they will tend to preserve some of their strength. Muraven, Shmueli, and Burkley (2006), in four experiments, tested the hypothesis that participants tend to conserve their strength for future acts of self-control. Their results showed that participants, who believed, that they will be involved in an additional future self-control task, performed more poorly in the critical task than participants who did not expect it. The experimenters explained their findings in light of the conservation hypothesis. However, the authors acknowledge that there may be a depletion level beyond which self-control exertion may no longer be possible and claim that ethical and practical limitations prevent testing of such a hypothesis in laboratory settings (Baumeister et al., 2007).

In a meta-analysis of 83 studies, Hagger et al. (2010) tested the predictions of the strength model of self-control and alternative explanations. Although their findings in general provided support for the ego depletion effect, their analysis did not confirm the hypothesis that continuous effort of self-control would lead to a gradual decline in performance. Studies that did not report an interim period between tasks found weaker ego-depletion effects than studies that reported an interim period, contradicting the replenishment hypothesis. Furthermore, the duration of the depleting task did not seem to affect the ego-depletion effect, failing to support the notion of a gradual decline in performance through continuous effort.

One possible reason for the absence of the effect in the aforementioned analysis is that the dual-task paradigm used in the self-control literature has a major

disadvantage, specifically, the use of two different self-control tasks. As most experimenters empirically know, interim periods in such experimental procedures are unavoidable: time is needed to start the task and give instructions for the second task to the participants. Although this time is not experimentally manipulated or calculated, it may affect the overall outcome. We can assume that there is plenty of time between the two tasks for the participants to restore some of the lost self-control strength and, probably, modify ego-depletion effects.

One frequently used task in the ego depletion literature is the Stroop task. For the ego depletion literature automatic and controlled responses are critical for the definition and examination of self-control, in general, and of self-control strength, in particular. Automatic responses are thought to be rigid, requiring more self-control strength to be inhibited, whereas inhibition of controlled responses incurs less cost in terms of self-control strength (Muraven & Baumeister, 2000). As Baumeister et al. (2007) proposed, the Stroop task, and more specifically the incongruent condition of the task, requires from the participants to override and inhibit a dominant and automatic response, namely word reading, something that can result in an ego-depletion state. The Stroop task can be either the depleting task or the manipulation check (dependent) task. One example of Stroop as the manipulation check task comes from the study of Gailliot, Schmeichel, and Baumeister (2006, Study 6), who were interested in the investigation of self-regulation in threat situation and depletion. In their sixth study participants in the depletion group were asked to imagine their death, whereas participants in the control group were asked to imagine an aversive topic. After that, both groups completed the Stroop task. The result showed that the depletion group performed more poorly on the Stroop task than the control group. In contrast, Neshast-Doost and Golden (2008) used the Stroop task as the depleting task in order to investigate the impact of depletion on the retrieval of emotion-related

autobiographical memories. So, the depletion group completed an incongruent condition of the Stroop task for 6 ½ min, whereas the control group read neutral color words (in black ink) or did nothing. After completion of the first phase, participants engaged in the Autobiographical Memory Test (AMT). AMT is a test that requires from the participants to retrieve specific autobiographical memories in 30 sec after seeing specific positive, negative, and neutral cue words. The results confirmed their hypothesis that depletion, caused by the incongruent condition of the Stroop task, may have an impact on autobiographical memory retrieval.

Although it is frequently used in the self-control studies as the depleting task, Hagger et al. (2010) found that the effect size for the modified Stroop task was smaller in magnitude compared to other tasks as for example the cross-out-letters task. One possible explanation may be that many studies adopt the single-item version of the Stroop task. Specifically, in the study of Govorun and Payne (2006), congruent and incongruent trial were mixed and presented individually on the screen. The depletion group completed 300 trials of the Stroop task (in approximately 15min) with breaks between blocks of 100 trials; the control group completed only 30 trials (approx. 1min). Participants in the study of Muraven, Rosman, and Gagné (2007) indicated the ink of only 80 words in the incongruent or congruent condition in a single-trial version. Pocheptsova, Amir, Dhar, and Baumeister (2009) used a much shorter version of the single-trial Stroop with only 40 trials and feedback between trials. Finally, Bruyneel, Dewitte, Franses, and Dekimpe (2009) used a short variation of the single-item version of the Stroop task. Participants had to name the ink of 50 color words. Color and word were either matched (congruent, control condition) or mismatched (incongruent, depletion condition). In addition, participants in the depletion group were asked to honor an exception: if the word appeared in green ink

(e.g., red), they had to indicate the word (say “red” and ignore the color). 25% of the total trials followed this exception.

The single trial version of the Stroop task usually contains standard stimulus presentation and interstimulus intervals. In some cases, like in the study of Pocheptsova et al. (2009), even feedback may interrupt the continuity of trials. This extra time given to the participants may allow them to rest between trials and replenish some of their self-control strength, leading, consequently, to underestimation of the ego-depletion effect. This becomes more plausible, if we take into account the short-term nature of depletion, underscored by its proposers. As Muraven and Baumeister (2000) state about their literature review indicating aftereffects of self-control, “...[the review] focuses on the crucial prediction that self-control is impaired when it follows *soon* after a previous self-control attempt” (p. 249, emphasis added). In fact, the central replenishment hypothesis of the model is based on the short-term nature of the phenomenon (Baumeister, 2002; Muraven & Baumeister, 2000).

In my view, it is of major importance, which test format of the Stroop task is chosen to be implemented in the various studies, not only considering the self-control strength model, but also the conflict monitoring hypothesis. This topic will be addressed in more detail in the next section.

Card versus Single-item Version

Tece and Dimartino (1965) were the first who developed the single-item version of the color-word Stroop task. In their experiment stimuli were presented in a tachistoscope. Each session contained 48 trials and each trial was presented individually until the participants responded.

Since this first modification of the traditional card version of the task, the single-item version has been widely used in cognitive and clinical settings. Since each item

appears individually on the screen, vocal or manual responses can be measured separately for each item in milliseconds. Total naming times for the entire task, as in the card version of the task, can be avoided. More importantly, errors can be excluded from the reaction time analysis and not confound the data, an inherent problem of the card version (MacLeod, 2000). The instructions to the participants are exactly the same in the two versions (“Respond as quickly as possible and try to avoid errors”). Salo, Henik, and Robertson (2001) state the basic differences between the two versions of the task: 1) the single-item version has no flanking stimuli that may influence and impair performance; 2) in the classical version of the task the conditions are blocked, whereas in the single-item version they are usually randomly presented within the same session; and 3) stimuli in the single-item version are presented at central fixation and no generation of eye-movements is needed. It should also be noted that in the single-item version, items are typically displayed for a fixed time and that there is a temporal gap between trials, namely the interstimulus interval, that may depend on the experimenter’s purposes (MacLeod, 2000).

Although Stroop interference is a robust phenomenon and is observed in both versions, its magnitude varies as a function of administration. As MacLeod (2000) points out, while in the single-trial version interference estimates—defined as the difference between the incongruent and control condition—are about 120 ms, in the card version they are 420ms. MacLeod (2000) proposes that this difference in interference is due to the speeded responses in the control condition of the card version (600 ms vs. 700 ms in the single-item version), because participants can look ahead and move more easily to the next stimulus, but also because nearby stimuli can produce more interference in the incongruent condition resulting in an increase of naming times (1,020 ms vs. 820 ms in the single-item version).

Because of the importance of these observations and the fact that both test formats are used interchangeably and compared between studies, Salo et al. (2001) took one step further and tried to examine the methodological differences between the two version of the Stroop task in three experiments. For all experiments a total of 36 stimuli in four colors (red, green, blue, and yellow) were used, and there were three repetitions of each block for each condition (72 responses per condition). For Experiments 1 and 3, the control condition consisted of repetitions of the letter X printed in one of the four colors (e.g., **XXX**). The incongruent condition followed the usual format (e.g., **green**) in all experiments. The dependent variable was response time per item. In the card version, stimuli were presented in four columns of nine items and participants were instructed to name the stimuli from top to bottom in columns. In Experiment 1, their purpose was to compare the three version of the Stroop task: the classical card version, the blocked single-trial version, where each condition is presented in different blocks, and the random single-trial version, where conditions are mixed randomly into the same block. Their results showed that the neutral condition of the card version was faster than that of the blocked and random single-trials versions (card: 491 ms/item; blocked: 553 ms/item; random: 587 ms/item). For the incongruent condition, card and random version response times were virtually identical (card: 680 ms/item; random: 681 ms/item) and the incongruent condition of the blocked version was the slowest (713 ms/item). Interference was greatest in the card version (189 ms) followed by the blocked version (160 ms). Interference in the random version was the lowest (94 ms). The authors concluded that the difference in interference between task versions is, mainly, due to differences in the control condition, which is facilitated in the card version. They raised the issue that the subtraction used to estimate interference, may not be an

adequate method, as it cannot be assumed that the control condition is stable across versions.

In light of the above findings, in their second experiment they examined the impact of various versions of the control condition. They hypothesized that the physical similarity of the stimuli (repeated XXX) in the blocked versions may lead to reduced reaction times compared to control conditions that use dissimilar stimuli. In this experiment they used three different control conditions: one with XXX only, one with physically dissimilar neutral stimuli (XXX, WWWW, SSSSS, MMMMM), and one with animal names (dog, bear, tiger, monkey). Administration was blocked in the card and single trial version. They found that the similar condition was faster than both dissimilar conditions, with the animal condition being the slowest (XXX: 624 ms; XSMW: 637 ms; Animals: 659 ms). Although the difference between the XXX and XSMW did not reach statistical significance, the authors drew the conclusion that physically similar neutral stimuli tend to decrease response times in the control condition and, consequently, increase interference.

Finally, in their third experiment, they examined attention shifting as an aspect of slowing in the card version because additional costs may ensue, due to attention shifts, to different locations of the card. In this experiment, only the single-trial versions (random, blocked) were implemented and stimuli in all conditions appeared sequentially. The first stimulus was presented in the upper left corner and the last stimulus ended up in the lower right corner, simulating the spatial location of stimuli in the card version. Surprisingly, the blocked version with location shift was faster than the random version in both conditions (neutral: 563 ms; incongruent: 706 ms, vs. neutral: 612 ms; incongruent: 726 ms, respectively). The blocked location-shift version and the blocked single-trial version comparison showed no statistically significant difference. In contrast, there was a significant difference between the

random location-shift version and the random single-trial version. According to the authors, the attentional shift to subsequent locations in the card version does not seem to influence performance.

Overall, Salo et al. (2001) suggested that the choice and implementation of different control conditions may be a critical factor for the differences observed between the two test formats. They summarized their main conclusion by claiming that "...the method and selection of neutral stimuli—that provide the baseline by which interference is measured—are critical because they clearly can change performance" (p. 462). However, the authors acknowledged that variations in performance may also occur due to variation in the reaction time calculation (summation or per-item estimation). Recall that in this study reaction time was measured per item, and errors were excluded from the analysis of only the single trial versions (as exclusion of errors in the card version can be difficult and time consuming).

In my view, the suggestion of Salo et al. (2001) and MacLeod (2000) that, in the card version, nearby stimuli speed up responses in the control condition, but interfere with the target stimulus in the incongruent condition, slowing responses down, seems contradictory. Why would nearby stimuli cause interference only in the incongruent condition and not in the control condition? Why would adjacent irrelevant items function as flankers only in the incongruent condition, when indeed there is evidence that color-color interference may exist, and flankers may influence performance, when occupying the same response channel.

Evidence for Adjacent Stimulus Interference in Both Conditions

The first piece of evidence in support of the hypothesis of adjacent stimulus interference in the card version of the Stroop task comes from studies examining the impact of integration of the two dimensions, namely word and color, on the

magnitude of interference. Although this is a critical issue, as noted by MacLeod (1991) in his review, it was (and still is) an understudied topic. However, the few existing studies give us a clue about how nearby stimuli may influence performance.

One of the first studies examining performance in a spatially separated version of the Stroop task is that of Gatti and Egeth (1978). Inspired by the Eriksen flanker task, the authors tested the hypothesis that participants process information in the entire task-dependent display and compared different distances between the target word and the flanker (1°, 2°, 3°, 5°). Distractors were compatible with the correct response (e.g., “red” in black ink above and below a red color patch), incompatible (e.g., “red” in black ink above and below a green color patch), or neutral (e.g., “XXX” in black ink above and below a green color patch). Conflicting distractors were found to interfere at all distances, although their impact was reduced as a function of distance.

A different experimental procedure was adopted in the study of Kahneman and Henik (as cited in Kahneman & Chajczyk, 1983), where two color words were again presented simultaneously at the same distance. In all conditions the target color word was inside a circle and the irrelevant color word was inside a square. The difference was that in one condition the color word was indeed colored, whereas in the other condition the color word was in black ink. Their results indicated that color word interfered with color naming (delay about 200 ms), whereas color words in black ink had little effect on color naming of the target color word. They proposed that attention to relevant dimensions of the task is reinforced, causing nontarget stimuli with relevant dimensions to interfere with the target.

In a further study, Kahneman and Chajczyk (1983) found that, when the participants' task was to name the color of the ink of a color bar in the presence of an incompatible color word, performance decreased. In contrast, when the irrelevant

word was a color-neutral word (e.g., **must**) conflict effects were diluted. The authors concluded that interference from irrelevant stimuli exists when they are task relevant.

More recently, Macleod (1998) conducted a study examining the effects of practice in the integrated and separated versions of the Stroop task. Participants practiced color naming for five days either in the traditional, integrated single-trial version of the Stroop task, or in a modified single-trial Stroop. In the latter version, colored asterisks appeared above a color word (in white ink), in the incongruent condition, or above XXX, in the control condition. Participants' task in the color naming condition was to identify the color of the asterisks. In this study, interference decreased with practice in both versions, although, as the authors noted, in the separated version interference was more resistant to change.

Because the main research goal of all the above studies was the examination of integration in the Stroop task, authors focused on word-color interference, i.e., the impact of a distracting word on color naming, and did not consider a possible color-color or word-word interference. Glaser and Glaser acknowledged these issues in two seminal papers in 1982 and 1989, and their findings are enlightening. Specifically, in 1982 they examined, among other things, word-word and color-color interference using different stimulus onset asynchronies (SOA). Participants were instructed to name/read the second color or word and ignore the first, or the opposite (name/read the first and ignore the second). The results showed that word-word interference is possible, and more interestingly, the same was true for color-color stimuli. In other words, a word distractor could interfere with the target word (maximum 107-ms delay), and a color distractor could interfere with the target color (maximum 87-ms delay). The authors proposed that when modally pure stimuli are used (color-color, word-word) interference always occurs. These findings were replicated in a second study (Glaser & Glaser, 1989, Experiment 2), where it was again found that color

naming suffered under the presence of an incongruent color (maximally 51-ms delay) and the same was true for word reading in the presence of a distracting incongruent word (maximally 113-ms delay). In my view, these findings suggest that in the card version of the Stroop task, where multiple target items are present, adjacent stimuli may interfere with the target and decrease performance, and that this could be true not only for the incongruent condition, but also for the control condition, where modally pure stimuli are simultaneously presented.

The notion of interference occurring in the control condition is further enhanced if we take into consideration findings coming from the Eriksen flanker task. In the original task, as proposed by Eriksen and Eriksen (1974), participants were instructed to respond to a central letter. When this letter was H or K they were asked to press a lever to the right, and when the letter was S or C one to the left. A target letter was presented in five different combinations: 1) a target letter was flanked by identical letters (e.g., HHHHHHH), 2) a target letter was flanked by a letter that corresponded to same motor response (e.g., KKKHKKK), 3) a target letter was flanked by a letter with incompatible response (e.g., SSSHSSS), 4) a target letter was flanked by a letter not included in the target set but sharing similar characteristics with the target (e.g., NWZHNWZ), and 5) a target letter was flanked by a letter not included in the target set and with dissimilar characteristics with the target (e.g., GJQHJQ). There was also one condition, in which the target letter appeared alone, without flankers. The distance between target and distractors was also manipulated. Three different spacings between target and distractor were implemented: 0.06° of visual angle, 0.5° , and 1° . The results showed that response-incompatible distractors produced increased response times compared to all other conditions, with maximal values in the closest distance. Also, dissimilar flanker letters that were similar to the incompatible response produced longer reaction times than when flankers were similar to the compatible

target set. The crucial observation in this experiment was that if flankers are members of the experimental design but bound with an incompatible response, they produce interference. If they are irrelevant to the experiment letters no such interference is produced. This finding supports a response competition hypothesis, where two responses relevant to the task are activated and competing, and the inappropriate one should be inhibited. The extra time needed to inhibit the irrelevant response results in increased reaction times in the incompatible condition.

In regard to the Stroop task, as a paradigm of response competition, Eriksen (1995) comments: “Our findings show that this response compatibility effect is much more general than had been assumed from experimentation with the Stroop task. The effect is not limited to conflict inherent within the stimulus itself; *it can come from other objects in the visual field in proximity to the attended object*” (p. 105, emphasis added).

Interestingly, the visual attentive field is not a priori defined, but depends on the task demands. It is not the absolute distance between target and distractor that determines the magnitude of interference but their distance from the attended area. If the task demands a larger field to be attended, then the distractors can interfere even if their distance from the target is large, but the distance from the attended area is small (Eriksen, Pan, & Bottela, 1993).

In the context of our research, all the above findings could mean that the card version of the Stroop task exhibits longer reaction times and greater interference than the single item version, not because the control condition is facilitated by the nearby stimuli, but because the sources of interference are more than one. If adjacent items belonging to the specific target set produce interference when incompatible with the current response, and if the attended field is enlarged due to task demands, we could infer two main hypotheses: a) there are more interference sources in the incongruent

condition (not only from the incongruent word, but also from the adjacent words and colors), and b) the control condition suffers from interference as well, since incompatible responses to the target are present (for an illustration of these assumptions see Figure 1.5, where items of the incongruent and control condition of the Stroop task are compared with a variant of the Eriksen flanker task).

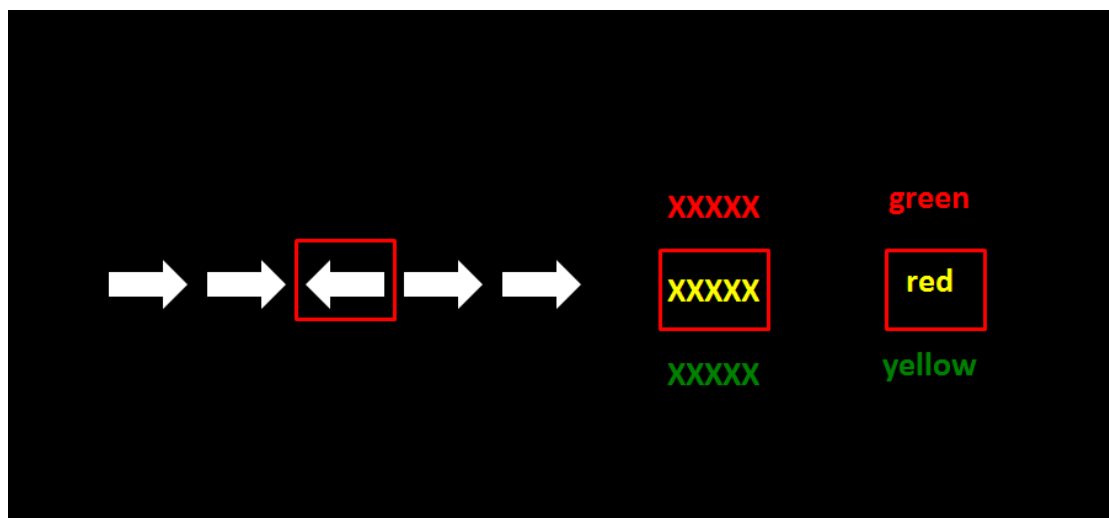


Figure 1.5. Comparison of a modified flanker task (left) to the card version Stroop task (right).

Rationale of the Present Study

In the present study our main goal was to test the hypotheses deriving from the conflict monitoring theory and the strength model of self-control within the same task, since they seem to lead to different conclusions.

According to the conflict monitoring hypothesis, a performance improvement should be observed in the incongruent condition of the card version of the Stroop task, because the condition is blocked (only incongruent items are present), and this means that the activation of the conflicting monitoring unit would be high during the entire task, leading to high activation of the control units and, ultimately, to performance improvements.

In contrast, according to the strength model of self-control, a gradual decline in performance should be observed, as the task progresses, and the participants become depleted.

We chose to adopt the card version of the Stroop task, and not the single-trial version, to test three additional assumptions. First, consider the fact that the card version of the task is extremely brief (approximately 102 seconds to name the colors of 100 stimuli in the incongruent condition; MacLeod, 2000). If ego depletion emerges during the task, this would mean that continuous effort and the absence of interim periods is the critical aspect for this phenomenon to emerge, rather than task duration.

Second, because the card version, due to its nature, is more complex than the single-trial version, if depletion is observed, this would mean that task complexity is a crucial moderator of ego depletion.

Third, and more importantly, we were interested in examining the course of the control condition, assuming that adjacent items interfere with the current responses. No such prediction can be derived from either of the above models. Specifically, according to the conflict monitoring model, as it presently stands, there is no source of conflict in the control condition, and the energy of the model should be equal to zero. Therefore, neither improvement nor impairment but, rather, stable performance should be observed. Otherwise, the model would need revision. Regarding, on the other hand, the strength model of self-control, no clear-cut hypothesis can be made. We can hypothesize that if interference from nearby stimuli exists, and if the task requires continuous effort, as we propose, then a gradual decline in performance should be evident in the control condition, too. If so, it would be of primary interest to examine how interference is influenced by it.

Experiment 1

Method

Participants

The total sample consisted of 41 adults 21-38 (32 female), mainly undergraduate and graduate students. All participants were native speakers of Greek.

Material and apparatus

The Greek words for red (κόκκινο /kocino/), green (πράσινο /prasino/), and yellow (κίτρινο /citrino/) were used, because they have the same number of letters and syllables and comparable written frequency (33, 34, and 9 per million, respectively, from the IPLR; Protopapas, Tzakosta, Chalamandris, and Tsiakoulis, 2012) . The corresponding colors are familiar and easily distinguishable.

Stimuli for the neutral color condition were made up of 7 repetitions of the letter X (no spaces) in red, green, and yellow color (RGB #FF0000, #00FF00, and #FFFF00, respectively). For the incongruent condition the Greek words for red, green, and yellow appeared in a non-matching color.

Each condition was presented in a single-screen array of three columns of 20 stimuli, resulting in a total of 60 stimuli per condition. Each item extended 19×94.5 pixels on the screen. The distance between items (edge to edge) was 22.7 pixels and between columns 124.70 pixels.

For the incongruent condition there were 20 repetitions of the word and 20 repetitions of the color, divided equally into the three columns. For the control condition 20 repetitions of each color were used, again equally divided into the three columns. The color words or colors were listed in random orders with the constraint that adjacent items were not the same.

Each card was presented on the screen of a 15.5'' laptop (resolution = 1280×1024 pixels) and color naming in each condition was recorded on the computer via a

headset using Speech Analyzer. The distance between adjacent stimuli, although not perfectly matched⁴ between conditions and during the task was approximately 0.5 degrees of visual angle at a distance of 50cm from the screen.

The stimuli for both conditions were in 20-pt Arial font displayed on a black background (*Appendix A*).

Procedure

The control condition was administered first followed by the incongruent. We asked the participants to name the color of the ink as quickly as possible and to try to avoid errors, moving from top to bottom in columns. Prior to the tasks the production of the intended colors and understanding of the instructions were verified by showing to the participants exemplar cards of the task. There were no practice trials.

Analyses

For both conditions, the total naming duration for each of the three columns for the two conditions was subsequently measured on the waveform using Praat (Boersma & Weenink, 2012). Mispronunciations and substitutions were considered errors. Two participants did not complete the task and were excluded, leaving 39 for analysis. Mean reaction times (RTs) were calculated for each subject for each condition and column and refer to item per column.

To examine the effects of column and condition on color naming times, errors (mean number of errors) and interference (incongruent minus control) we used function lmer of the lme4 package (Bates, Maechler, & Bolker, 2012) in R (R Core Team, 2012). We employed linear mixed-effects, that is, including random intercepts for participants as well as by-participants random slopes for column and condition. For response time, in R notation the model formula was specified as:

⁴Administration of the Stroop task does not restrict participants to maintain a specific behavior during the task, e.g. "do not approach the screen". If any such restrictions would be imposed here, our study would lose its comparability power. However, it should be noted that the vast majority of the participants remained at the indicated distance. This stands for all Experiments described.

RT ~ column*condition + (1+column+condition|subject)

For this model, “column” refers to the first versus second versus third column, and “condition” to incongruent versus control. Independent variables were deviation coded. The significance of the interaction was tested with comparison to a simpler, without interaction model.

Posthoc pairwise analyses with Bonferroni correction followed interaction analyses using function lme of nlme package (Pinheiro et al., 2014)

Results

Results showed that the column and condition affected color naming. Color naming time decreased as the task progressed (first column vs. third $\beta = -0.07$, $t = -9.29$; second column vs. third, $\beta = 0.03$, $t = 4.74$), and the control condition differed from the incongruent ($\beta = 0.10$, $t = 13.40$). Also the interaction column \times condition was significant ($\chi^2 = 15.82$, $df = 2$, $p < 0.001$; Table 1.1).

Post hoc analysis with Bonferroni correction for three comparisons in each condition indicated that, in the incongruent condition, the first column differed from the second ($\beta = 0.14$, $z = 6.302$, $p < 0.001$) and from the third ($\beta = 0.14$, $z = 6.245$, $p < 0.001$), but there was no difference between the second and the third column ($\beta = -0.001$, $z = -0.058$, $p > 0.05$). In the control condition, the first column differed from the second ($\beta = 0.06$, $z = 6.620$, $p < 0.001$) and the third ($\beta = 0.09$, $z = 9.232$, $p < 0.001$). In the control condition the second column differed also from the third ($\beta = 0.02$, $z = 2.611$, $p < 0.05$; Figure 1.6).

Table 1.1.

Color naming times and number of errors per column in each condition

Column	Naming Times(s)				Number of errors			
	Incongruent		Control		Incongruent		Control	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1	0.69	0.11	0.52	0.08	0.01	0.03	0.00	0.01
2	0.83	0.13	0.58	0.09	0.02	0.04	0.00	0.01
3	0.83	0.16	0.61	0.10	0.04	0.02	0.41	0.02

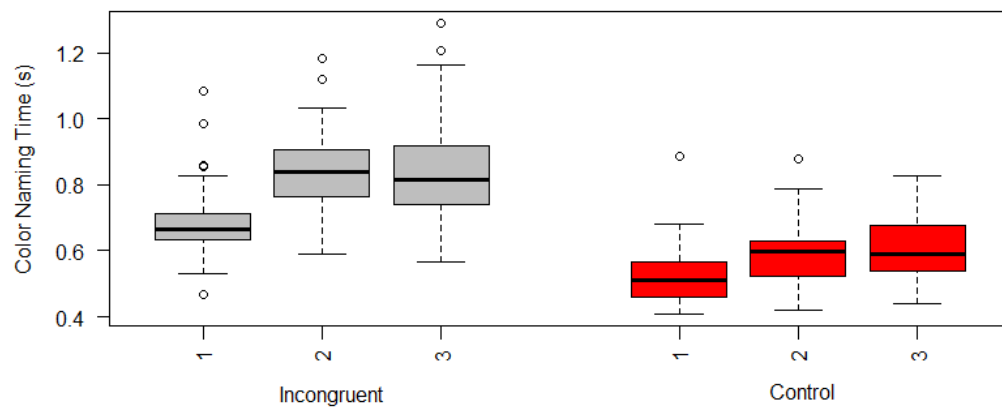


Figure 1.6. Color naming time per column in each condition.

Stroop interference was calculated as the difference in color naming time between the incongruent and the control condition in each of the three columns and was analyzed with linear mixed models using function lme.

The results indicated that interference differed between columns ($F(2,76) = 5.70$, $p < 0.01$). Post hoc analysis with Bonferroni correction revealed that the first column differed from the second ($\beta = 0.08$, $z = 3.321$, $p < 0.01$) but there was no difference between the first and the third column ($\beta = 0.05$, $z = 2.201$, $p = 0.08$), or between the second and third column ($\beta = -0.02$, $z = -1.120$, *ns*; Figure 1.7).

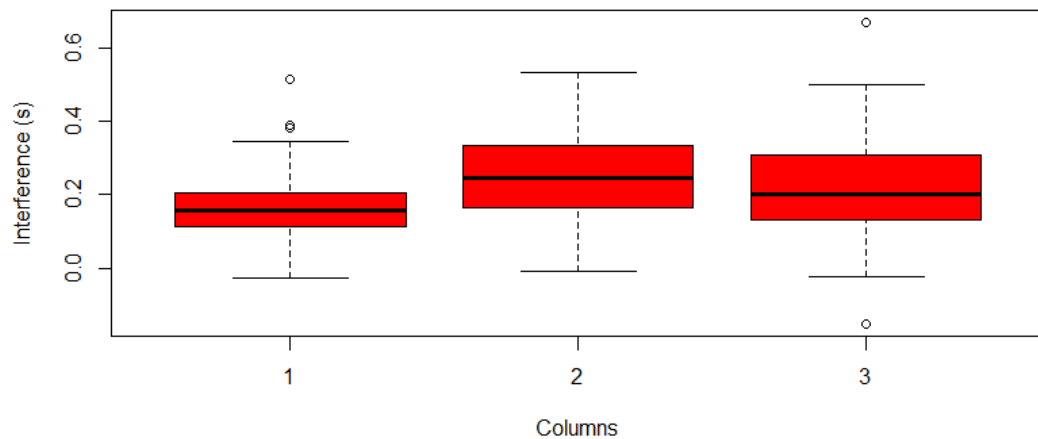


Figure 1.7. Interference in each column.

However, because participants named the control condition more slowly as the task progressed, perhaps the two last columns do not constitute an adequate baseline for interference estimation (cf. Dulaney & Rogers, 1994). Therefore, we recalculated interference in each column using as a baseline only the first column of the control condition. The employment of lme revealed again that interference differed between columns ($F(2,76) = 26.23, p < 0.001$). Post hoc comparisons with Bonferroni correction showed that interference differed between the first and second column ($\beta = 0.14, z = 6.302, p < 0.001$), but not between second and third column ($\beta = -0.001, z = -0.058, ns$). However, there was now a significant difference between the first and the third column ($\beta = 0.14, z = 6.245, p < 0.001$), suggesting that changes within the control condition may affect estimates of interference (Figure 1.8).

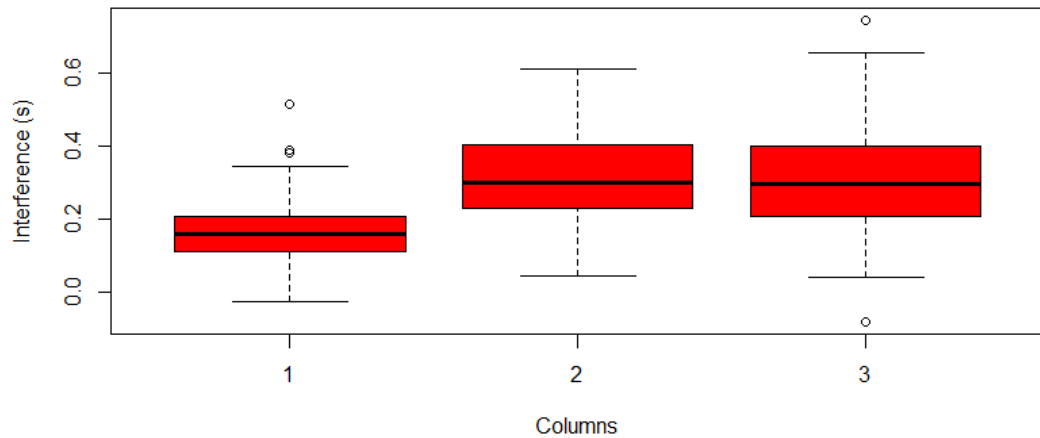


Figure 1.8. Interference in each column keeping the baseline constant.

Finally, for accuracy data (number of errors) we employed a general linear mixed-effects model. The results showed that there was a difference in errors between the first and third column ($\beta = -0.008$, $t = -2.67$), but there was no difference between second and third column ($\beta = -0.003$, $t = -1.17$). There was a difference between the control and the incongruent condition ($\beta = 0.008$, $t = 3.72$), but no interaction ($\chi^2 = 0.30$, $df = 2$, ns ; Figure 1.9).

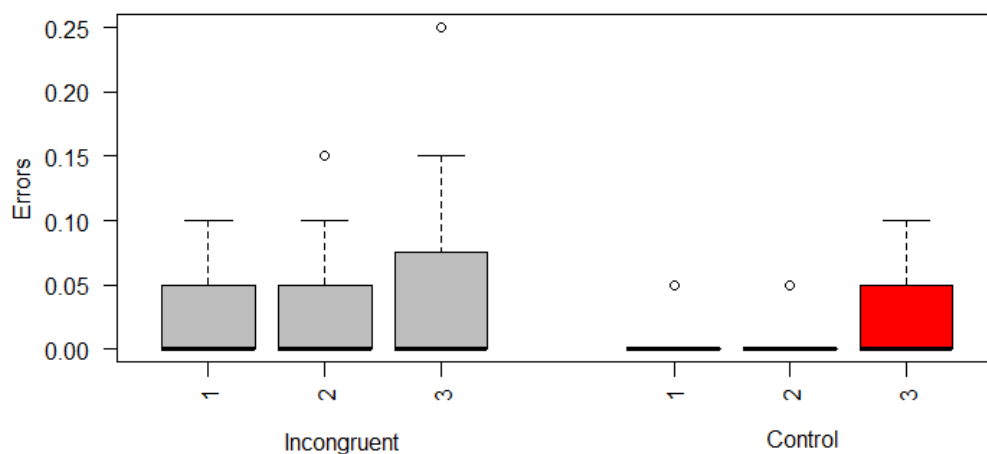


Figure 1.9. Mean number of errors per column in each condition.

For summary of the main findings in Experiment 1, see Box 1.1.

Box 1.1.*Main findings in each condition*

Incongruent Condition
Color Naming
<ul style="list-style-type: none"> • <i>Difference between first and second / third column</i>
Errors
<ul style="list-style-type: none"> • <i>Difference between first and third column (increase)</i>
Control Condition
Color Naming
<ul style="list-style-type: none"> • <i>Difference between first and second / third column</i> • <i>Difference between second and third column</i>
Errors
<ul style="list-style-type: none"> • <i>Difference between first and third column (increase)</i>
Interference
<ul style="list-style-type: none"> • <i>Difference between first and second column</i>

Discussion

Our findings contradict the conflict monitoring hypothesis and favor the strength model of self-control. First, in the incongruent condition there was no performance improvement, as the model of Botvinick et al. (2001) suggests. In contrast, performance impairment was evident, as was expected by the strength model of self-control.

Moreover and importantly, this decrement in performance was not limited to the incongruent condition, but also extended to the control condition. There is no possible way that the conflict monitoring hypothesis and the proposed model could account for this finding, unless revisited. Specifically, according to model of conflict monitoring, there is no reason in the control condition for conflict to occur: there is only one kind of response, namely color naming, and consequently, no alternative response to compete with it.

In contrast, these findings are in line with our proposal that in the card version of the Stroop task there is more than one source of interference and is present in both conditions. In the incongruent condition, the source is the incongruent word of the target stimulus and, in addition, the adjacent words and colors, whereas in the control condition the source is the nearby stimuli, included in the target set of the task with mutually excluding responses. Response competition appears in both conditions; it is the number of competitors that changes between them.

To the extent that there was no significant difference in errors between the first and second column in our experiment, posterror slowing as an alternative can be at the moment excluded. Posterror slowing refers to the observation that participants tend to be more “conservative” and slow down after errors (Carter & van Veen, 2007). For this to constitute an alternative explanation for the increase in color naming times between columns, more errors should have been observed in the second column. This was not the case. Nor can a speed-accuracy tradeoff explain our data, since participants slowed down but made the same number of errors.

Although the evidence of Experiment 1 seems to suggest multiple interference in both conditions of the card version of the Stroop task no final conclusions can be drawn, before we compare this version of the task with its single-trial counterpart. If not, one possible objection to our claims would be that it is not adjacent stimuli that impair performance but the number of items. In Experiments 2A and 2B, we tried to investigate both alternative explanations.

Experiments 2A and 2B

The aim of Experiments 2A and 2B was to compare the course of color naming in the card version and the single trial version of the Stroop task. If the decline in performance observed in Experiment 1 was, due to the interference caused by nearby stimuli, as we suggest, then no such decline in performance should be observed in the

single trial version. In contrast, if this decline emerges, due to other factors such as the number of items, then the same pattern of results should be observed in both versions.

As we were primarily interested for the course of the control condition because of its importance in interference estimation, one alternative to our suggestion was also examined. Could it be that the similarity of the items (all items were colored XXXs) contributed to the pattern observed? Treisman & Gelade (1980) have shown that the increase of RTs is possible, if the target and the flankers share common features like shape and color. Additional evidence derive from studies examining the crowding effect. Crowding refers to impaired identification of a target objects when flanked by other objects in close distance (Bouma, 1970), and there are findings that support the notion that similarity between target and flankers enhances the phenomenon, where dissimilarity breaks it (Whitney & Levi, 2011). Crowding is proposed as an alternative to Eriksen flanker task (Dayan & Solomon, 2010), so it would be reasonable to assume that if crowding stands for this task, then it could stand for the card version of the Stroop task as well.

Also, an observation of crowding studies is the substitution phenomenon: participants tend to mistakenly report a flanker instead of the target (Whitney & Levi, 2011). This phenomenon is also reported for the Eriksen flanker task (Eriksen, 1995), and is thought to be an additional index for interference caused by nearby stimuli. If this kind of interference exists in the card version of the Stroop task, then errors would be false reports of nearby items. In addition, if similarity of the items in the control condition can partially account for the gradually decline in performance, this decline will be attenuated as the dissimilarity increases.

Moreover, it has been observed that different control conditions are used between studies (MacLeod, 2000), and, as Salo et al. (2001) propose, it could be that the difference of interference estimates between studies is, due to the implementation of

various control conditions. So, it was of vital importance to investigate if the gradually decline in performance would be present in different control conditions, and to what extent, and how this decline could affect interference scores. Any analysis regarding the control condition without such a manipulation, would be inconclusive.

An additional goal of Experiments 2A and 2B was to examine the gradual decline in performance in the Stroop task, and its impact in interference in different populations. It is well established that interference shows a U-shaped function that begins as the individual learns to read, and reaches its highest level in Grades 2 and 3, and then decreases (MacLeod, 1991). So, one critical question is, if children will show a comparable to adults' pattern of results, or not.

Experiment 2A

Method

Participants

The total sample consisted of 20 adults 21-36 years old (12 female), mainly undergraduate and graduate students. All participants were native speakers of Greek.

Material and apparatus

As in Experiment 1, the Greek words for red, green, and yellow were used. Also, the incongruent and the control condition (XXX) with the same criteria were implemented. The difference was the addition of two more control conditions.

The XBL condition (*Appendix A*) was constructed by seven repetitions of the Greek letters X or B or L in three different colors (e.g., **XXXXXXXX**, **BBBBBBB**, **LLLLLLL**). These three letters were chosen because they do not share the same starting letter with the target colors used, and result in a similar length with the XXX condition, if put together as units (19×94.50 pixels).

The ANI condition (*Appendix A*) was constructed by the Greek words for whale (φάλαινα /falena/), gorilla (γορίλας /gorilas/), and deer (ζαρκάδι, /zarkaði/). These

animal names were chosen on the basis of three criteria: a) initial letter does not coincide with the initial letter of the target color, b) same number of letters and syllables with target colors, c) comparable written frequency (0.24, 0.38, 0.16 per million, respectively, from the IPLR; Protopapas et al., 2012)⁵.

Both versions in both tasks satisfied the constraint criteria of Experiment 1, regarding repetitions of words and colors in the three columns, and distance between stimuli.

The single trial task consisted of 60 stimuli in 20-pt Arial font displayed randomly in a black background (*Appendix A*). Each stimulus appeared on the screen for 2s. The interim period between stimuli was 167.67 ms. Each single-item task lasted approximately 3min.

All stimuli used in both tasks had the same length (19×94.50 pixels). The visual angle for the same distance was as in Experiment 1.

Procedure

Procedure was the same as in Experiment 1, with the exception that presentation of the conditions was randomly defined across participants. Half of the subjects saw the single-trial task first, and the other half second. There were no practice trials in both conditions. For the single-trial task three representative stimuli from each condition were presented to the participants for verification of responses and understanding of instructions.

Analyses

As in Experiment 1, the total naming duration for each of the three columns for the four conditions was subsequently measured on the waveform using Praat (Boersma & Weenink, 2012). For single trial tasks responses were examined with CheckVocal (Protopapas, 2007) to determine accuracy and placement of the timing

⁵Although animal names and color words differ regarding written frequency, it was not possible to obtain animal names with comparable written frequency to color words without sharing the same initial letter.

marks. Because data processing of the latter was in ms, response times for the single-item versions were transformed to sec, in order to be comparable with data obtained from the card version.

Mean RTs were calculated for each subject for each condition, column, and task resulting to an item per column estimation.

Mispronunciations and substitutions were defined as errors. Omissions in the card version and no responses in the single trial version were not included in the accuracy analysis. Mean number of errors was calculated for accuracy analysis. In the single-item version, errors were excluded from the analysis of response times. For the card version this was not possible, and errors were included (to be comparable with other studies; e.g., Salo et al., 2001).

There were mainly three interference estimations for each control condition by subtracting each control condition and each column from the incongruent corresponding column (e.g., INC/column1 minus XXX/column1 or ANI/column1 etc., for all conditions and columns).

We employed linear mixed-effects model analyses for color naming time testing the triple interaction between column, condition, and task. We included random intercepts for participants, and by-participants random slopes for condition and task. By-participants random slopes for column did not improve the models and were excluded. For color naming time the formula was specified as:

$$RT \sim \text{column} * \text{condition} * \text{task} + (1 + \text{condition} + \text{task} \mid \text{subject})$$

Here, “column” refers to column1 vs. column2 vs. column3, “condition” to INC vs. XXX vs. ANI vs. XBL, and “task” to card vs. single.

Because the above model did not converge with the above random effects for errors and interference, we implemented simpler models.

For the significance of the triple interaction the full model was compared to a simpler model excluding the interaction.

To follow up the triple interaction simpler models were employed with all possible combinations of fixed effects (i.e. condition× task, column× task, condition× column). Because simpler models for response times did not converge with all the above random effects included, only random intercepts by participants, and by-participants random slopes for task were included (selection after fit comparison).

Posthoc pairwise analyses with Bonferroni correction followed interaction analyses.

Results

Table 1.2 shows the naming times per column for each condition and each task.

Results showed that the triple interaction was significant indicating that response times varied as a function of column, condition, and task ($\chi^2 = 114.43$, $df = 17$, $p < .001$). Subsequent interaction analyses with simpler models indicated that that the interaction for condition and task, and for task and column was significant ($\chi^2 = 17.87$, $df = 3$, $p < .001$; $\chi^2 = 33.45$, $df = 2$, $p < .001$, respectively). Column and condition did not interact significantly ($\chi^2 = 6.46$, $df = 6$, *ns*).

Table 1.2

Color naming times and mean number of errors per column in each condition and each task.

Column	Naming Times(s)				Number of errors				
	Card		Single		Card		Single		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
INC									
1	0.69	0.17	0.63	0.10	0.03	0.05	0.01	0.04	
2	0.83	0.23	0.62	0.10	0.03	0.03	0.01	0.02	
3	0.80	0.23	0.62	0.12	0.03	0.06	0.00	0.01	
XXX									
1	0.56	0.14	0.53	0.08	0.00	0.01	0.00	0.01	
2	0.64	0.18	0.53	0.09	0.00	0.01	0.00	0.02	
3	0.67	0.14	0.54	0.08	0.01	0.02	0.00	0.01	
ANI									
1	0.57	0.15	0.58	0.07	0.00	0.00	0.00	0.01	
2	0.67	0.15	0.57	0.08	0.01	0.02	0.00	0.01	
3	0.73	0.22	0.58	0.09	0.02	0.04	0.00	0.01	
XBL									
1	0.59	0.15	0.55	0.08	0.00	0.01	0.00	0.01	
2	0.63	0.13	0.56	0.08	0.01	0.03	0.00	0.01	
3	0.65	0.12	0.56	0.09	0.01	0.02	0.00	0.02	

Note: “INC”: incongruent condition; “XXX”: control condition with repetition of X; “ANI”: control condition with animal names; “XBL”: control condition with repetition of X or B or L.

Post hoc analysis with Bonferroni correction for three comparisons in each condition in each task was employed. In the single trial INC condition, there was no statistical difference between columns (all β s > -0.01 , $p > .5$). In the single trial XXX condition again, there was no statistical significant difference between columns (all β s > 0.006 , $p = 1$). The same results were obtained for single trial ANI condition (column1 vs. column2, $\beta = -0.01$, $z = -1.17$, *ns*; column1 vs. column3, $\beta = -0.005$, $z = -0.61$, *ns*; column2 vs. column3, $\beta = 0.004$, $z = 0.56$, *ns*), and for the single trial XBL condition (all β s > 0.01 , $p > .2$; Figure 1.10).

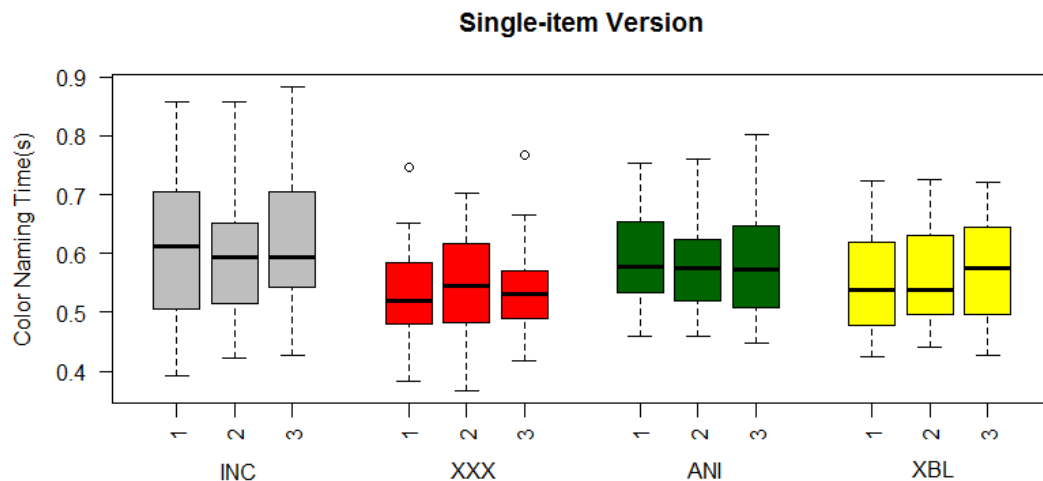


Figure 1.10. Color naming time in the single-item version per column in each condition.

In contrast, in the card version in INC condition, the first column differed from the second ($\beta = 0.13$, $z = 5.80$, $p < .001$) and from the third ($\beta = 0.10$, $z = 4.56$, $p < .001$), but there was no statistical difference between the second and third column ($\beta = -0.02$, $z = -1.23$, $p > .05$). In the XXX and ANI condition, the same pattern of results was observed: the first column differed from the second (XXX: $\beta = 0.07$, $z = 4.09$, $p < .001$; ANI: $\beta = 0.09$, $z = 2.92$, $p < .05$) and from the third (XXX: $\beta = 0.10$, $z = 5.59$, $p < .001$; ANI: $\beta = 0.15$, $z = 4.59$, $p < .001$). Second and third column did not differ in both conditions (XXX: $\beta = 0.02$, $z = 1.50$, *ns*; ANI: $\beta = 0.05$, $z = 1.66$, *ns*). In the XBL condition, a different pattern was obtained: there was no significant

difference between first and the second column ($\beta = 0.03$, $z = 2.12$, *ns*) or the second and third ($\beta = 0.06$, $z = 2.31$, *ns*). The first and third column differed ($\beta = 0.06$, $z = 4.43$, $p < .001$; Figure 1.11).

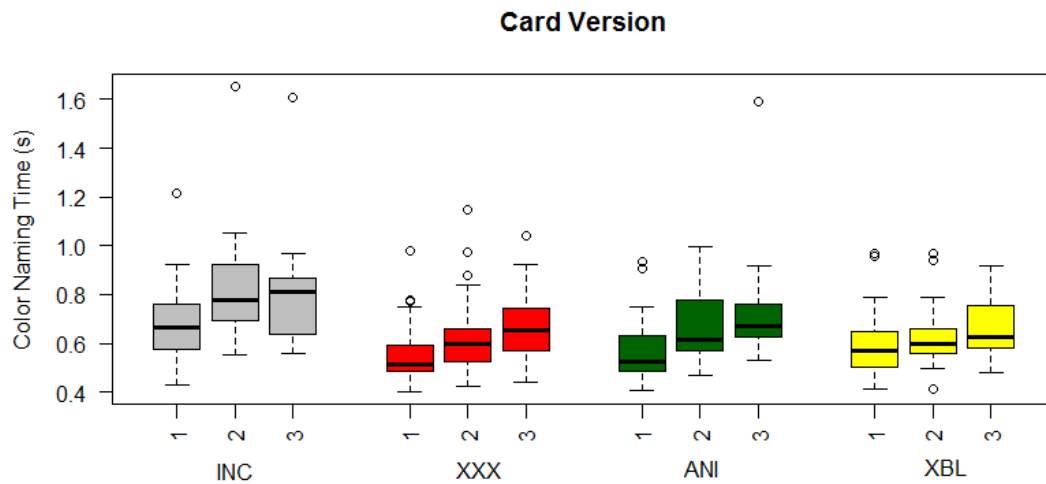


Figure 1.11. Color naming time in the card version in each condition.

For interference, the triple interaction $\text{column} \times \text{condition} \times \text{task}$ was significant ($\chi^2 = 40.63$, $df = 12$, $p < .001$). Simpler models revealed that the only significant interaction, was that of column and task ($\chi^2 = 18.51$, $df = 12$, $p < .001$).

Separate posthoc pairwise comparisons with Bonferroni correction for interference for each condition and task were performed. For the single trial task, interference for the XXX condition did not differ between columns (all $\beta_s > -0.01$, $p > .5$). The same pattern was observed for the single trial ANI condition (all $\beta_s > -0.01$, $p = 1$) and XBL condition (all $\beta_s > -0.01$, $p > .5$; Figures 1.12–1.14).

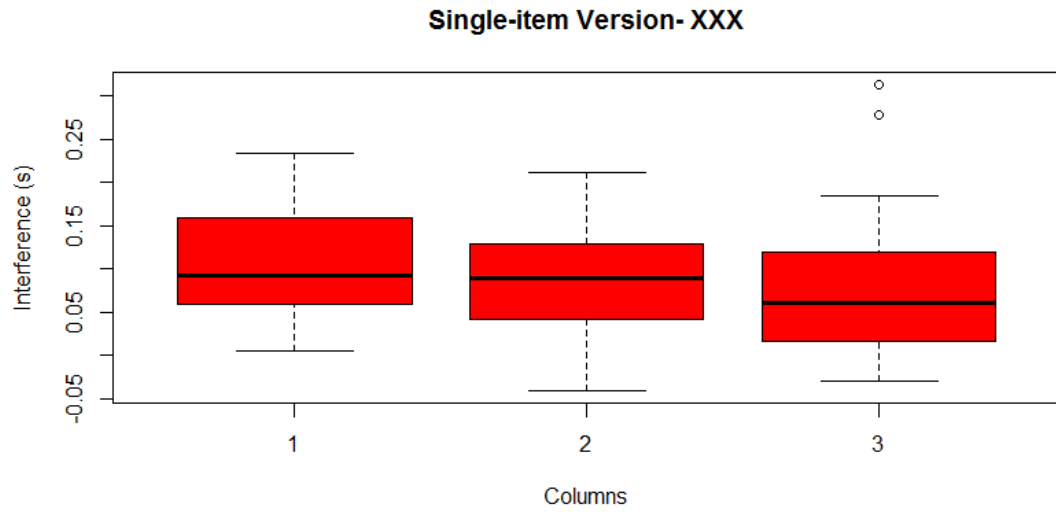


Figure 1.12. Interference in the single-item version using as baseline the XXX condition.

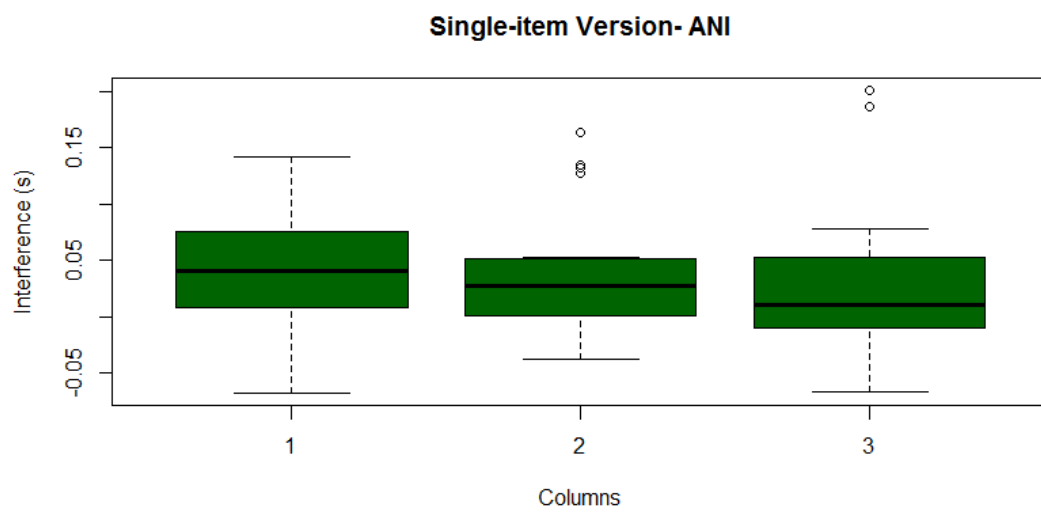


Figure 1.13. Interference in the single-item version using as baseline the ANI condition.

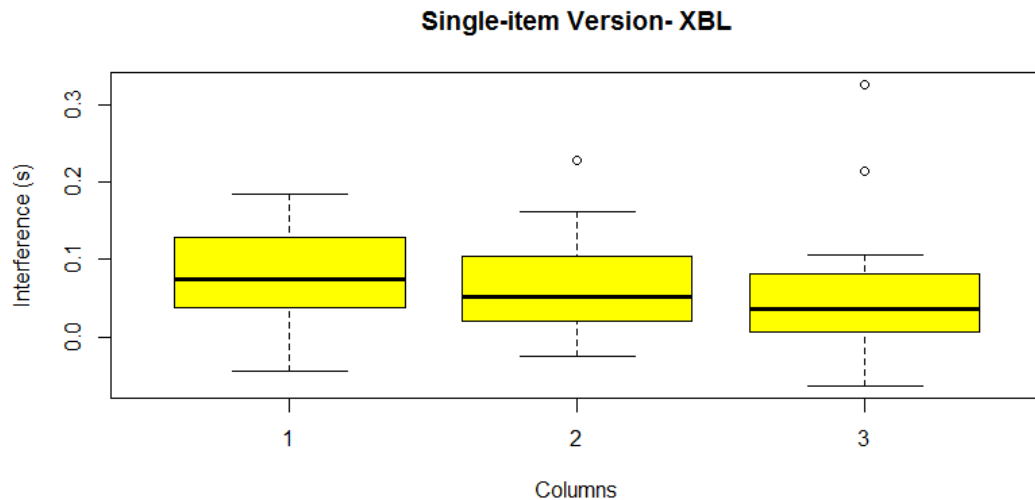


Figure 1.14. Interference in the single-item version using as baseline the XBL condition.

In contrast in the card version, for the XXX condition there was significant difference in interference between the first and the second column indicating increase, and the second and third indicating decrease (column1 vs. column2, $\beta = 0.06$, $z = 2.69$, $p < .05$; column2 vs. column3, $\beta = -0.05$, $z = -2.53$, $p < .05$). There was no difference between the first and third column (column1 vs. column3, $\beta = 0.003$, $z = 0.168$, *ns*). For the ANI condition no difference between columns in interference was observed, although for second and third this was marginal insignificant (column1 vs. column2, $\beta = 0.03$, $z = 1.07$, *ns*; column1 vs. column3, $\beta = -0.04$, $z = -1.25$, *ns*; column2 vs. column3, $\beta = -0.08$, $z = -2.32$, $p = .052$). For the XBL condition, interference was statistical different between the first and the second column (column1 vs. column2, $\beta = 0.10$, $z = 3.95$, $p < .001$), but not between the first and third column, or the second and third (column1 vs. column3, $\beta = -0.04$, $z = -1.59$, *ns*; column2 vs. column3, $\beta = -0.06$, $z = -2.36$, *ns*; Figures 1.15–1.17).

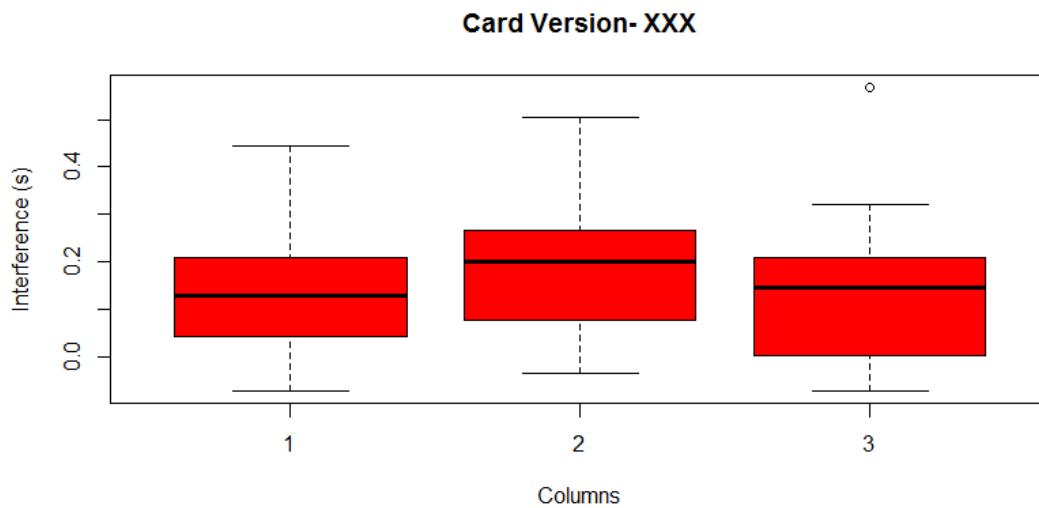


Figure 1.15. Interference in the card version using as baseline the XXX condition.

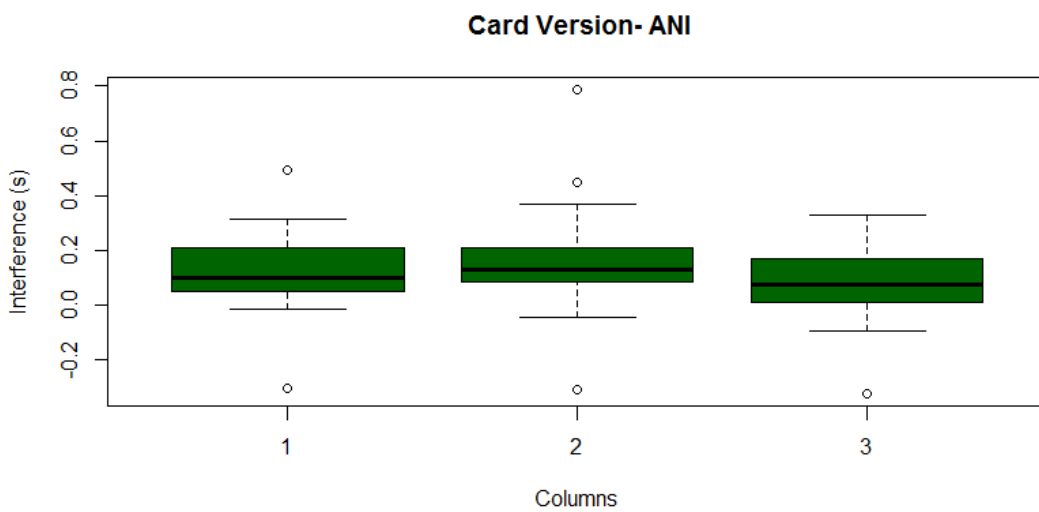


Figure 1.16. Interference in the card version using as baseline the ANI condition.

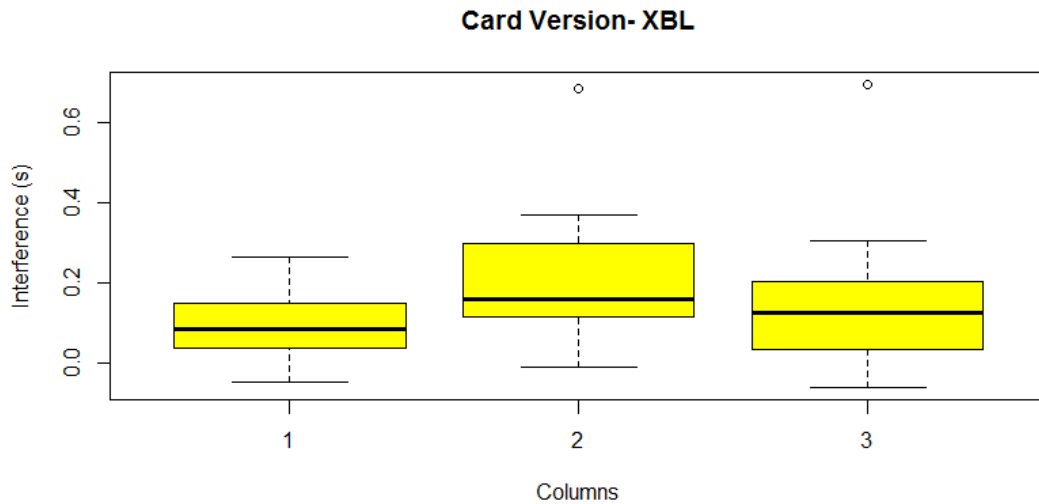


Figure 1.17. Interference in the card version using as baseline the XBL condition.

Since, as in Experiment 1, color naming time gradually increased in the control conditions of the card version, we recalculated interference for that version in each column in each condition, using as the baseline only the first column. The interaction of column, condition, and task was significant ($\chi^2 = 124.38$, $df = 4$, $p < .001$). Simpler models revealed that the interaction of condition and task, and for column and task was significant ($\chi^2 = 9.04$, $df = 2$, $p < .05$; $\chi^2 = 73.15$, $df = 2$, $p < .001$; respectively). The interaction of column and condition was nonsignificant ($\chi^2 = 0$, $df = 4$, ns).

Posthoc pairwise comparisons with Bonferroni correction showed that for the XXX condition, there was now a significant difference between the first and second column, and the first and third indicating an increase (column1 vs. column2, $\beta = 0.13$, $z = 5.80$, $p < .001$; column1 vs. column3, $\beta = 0.10$, $z = 4.56$, $p < .001$). The difference between the second and third column was not significant (column2 vs. column3, $\beta = -0.2$, $z = -1.23$, ns). The exact same pattern was observed for the ANI condition (column1 vs. column2, $\beta = 0.13$, $z = 5.80$, $p < .001$; column1 vs. column3, $\beta = 0.10$, $z = 4.56$, $p < .001$; column2 vs. column3, $\beta = -0.2$, $z = -1.23$, ns), and for the XBL condition (column1 vs. column2, $\beta = 0.13$, $z = 5.80$, $p < .001$; column1 vs. column3,

$\beta = 0.10$, $z = 4.56$, $p < .001$; column2 vs. column3, $\beta = -0.2$, $z = -1.23$, *ns*; Figures 1.18–1.20).

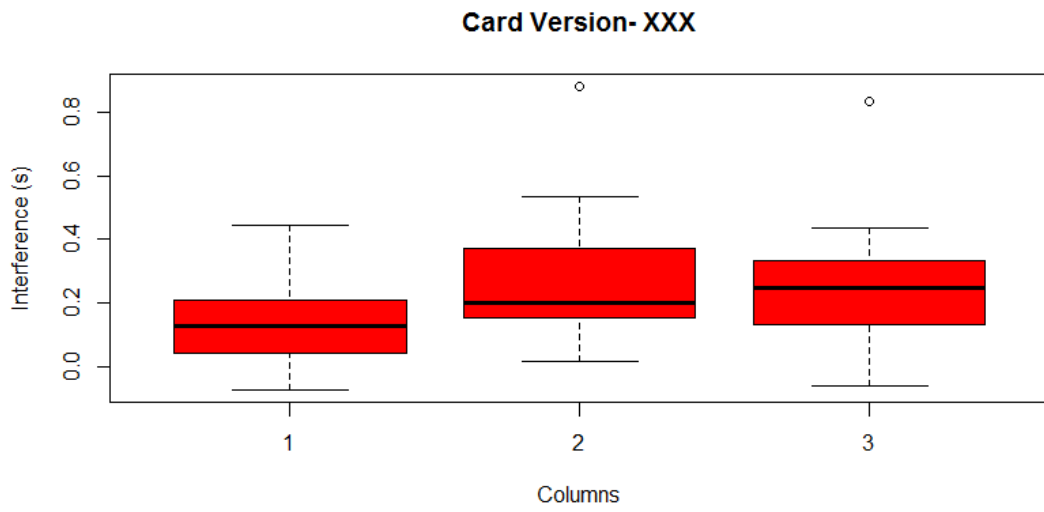


Figure 1.18. Interference in the card version keeping XXX baseline constant.

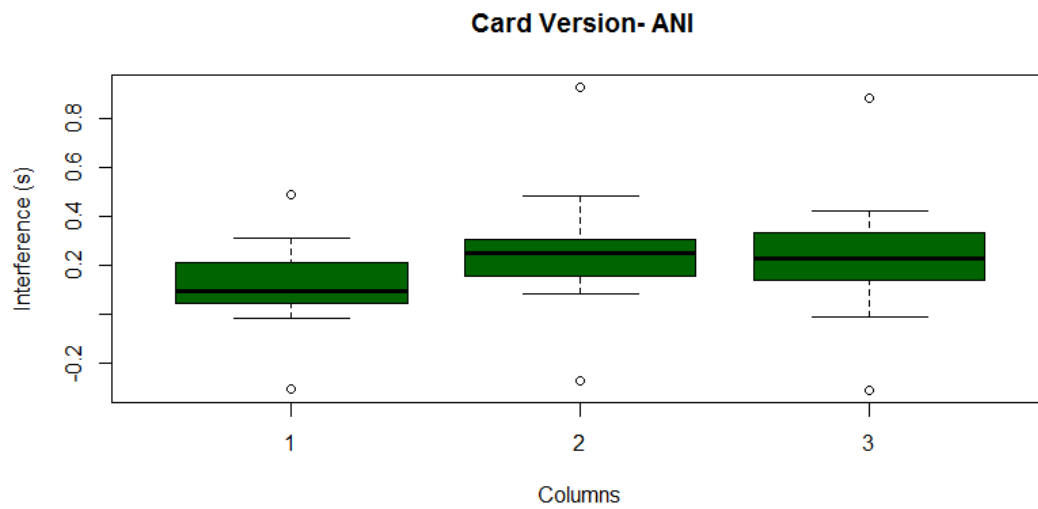


Figure 1.19. Interference in the card version keeping ANI baseline constant.

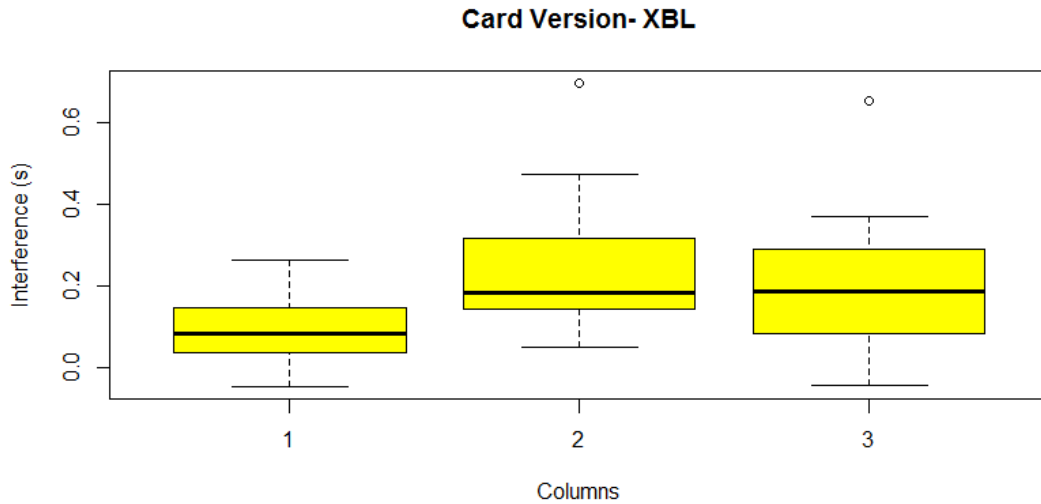


Figure 1.20. Interference in the card version keeping XBL baseline constant.

For accuracy data (mean number of errors) we employed a generalized linear mixed-effects model. The results showed that the triple interactions was nonsignificant ($\chi^2 = 25.89$, $df = 17$, ns). Subsequent analysis indicated that column and task interacted significantly ($\chi^2 = 9.28$, $df = 2$, $p < .001$) and the same was observed for task and condition ($\chi^2 = 8.67$, $df = 3$, $p < .05$), but not for column and condition ($\chi^2 = 4.37$, $df = 6$, ns).

Posthoc pairwise comparisons with Bonferroni correction for the single item version, revealed that columns did not differ for the incongruent condition (all β s < -0.01 , $p > .25$). The same was observed for the XXX condition (column1 vs. column2, $\beta = 0.005$, $z = 0.80ns$; column1 vs. column3, $\beta = 0.000$, $z = 0.00$, ns ; column2 vs. column3, $\beta = -0.005$, $z = -1.10$, ns), the ANI condition (column1 vs. column2, $\beta = -0.002$, $z = -0.43$, ns ; column1 vs. column3, $\beta = 0.000$, $z = 0.00$, ns ; column2 vs. column3, $\beta = 0.002$, $z = 0.43$, ns), and the XBL condition (all β s < -0.000 , $p = 1$; Figure 1.21).

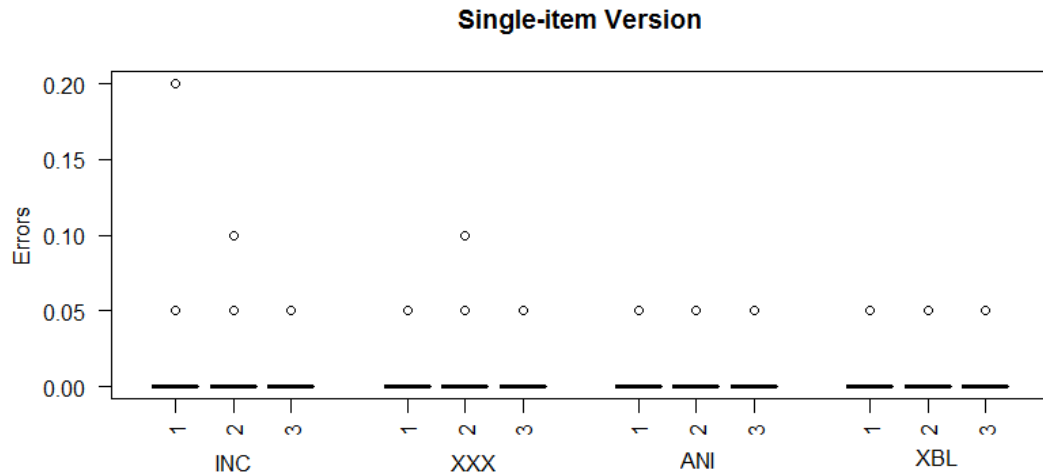


Figure 1.21. Mean number of errors in the single-item version per column in each condition.

For the card version, the mean number of errors did not differ between columns for INC condition (column1 vs. column2, $\beta = -0.000$, $z = 0.00$, *ns*; column1 vs. column3, $\beta = 0.007$, $z = 0.68$, *ns*; column2 vs. column3, $\beta = 0.007$, $z = 0.68$, *ns*). For the XXX condition, the first and second column, and the second and third did not differ (column1 vs. column2, $\beta = 0.002$, $z = 0.44$, *ns*; column2 vs. column3, $\beta = 0.01$, $z = 2.20$, *ns*), but there was a significant difference between the first and third column ($\beta = 0.01$, $z = 2.65$, $p < .05$). The same was observed for the ANI condition (column1 vs. column2, $\beta = 0.01$, $z = 1.18$, *ns*; column2 vs. column3, $\beta = 0.01$, $z = 1.78$, *ns*; column1 vs. column3, $\beta = 0.02$, $z = 2.96$, $p < .05$), but not for the XBL condition (column1 vs. column2, $\beta = 0.01$, $z = 1.72$, *ns*; column2 vs. column3, $\beta = -0.002$, $z = -0.345$, *ns*; column1 vs. column3, $\beta = 0.01$, $z = 1.38$, *ns*; Figure 1.22).

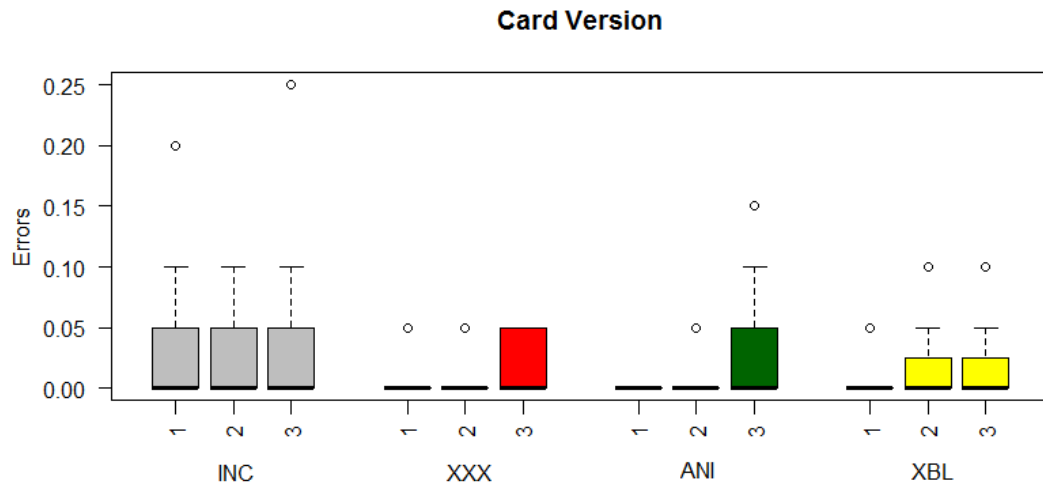


Figure 1.22. Mean number of errors in the card version per column in each condition.

Substitution errors were defined as errors deriving from the downward nearby stimuli of the target, and were calculated for all three control conditions⁶. For the XXX condition, substitution errors were 20% of the total errors, for the ANI condition 60%, and for the XBL condition 62.5%. For the ANI condition, we also calculated errors involving the word of the target (reading errors), but these were zero for adults. For a summary of the main findings in Experiment 2A, see Box 1.2.

⁶ Substitution errors did not include errors deriving from items above the target, because we could not infer with certainty, if they were indeed substitutions or repetitions of the previous stimulus.

Box 1.2

Main findings in each condition and task.

Single-item version

Color naming

- *No time difference between columns in all conditions*

Errors

- *No difference between columns in all conditions*

Interference

- *No difference between columns in all conditions*
-

Card version

Color Naming

- *INC condition: difference between first and second / third column*
- *XXX condition: difference between first and second / third column*
- *ANI condition: difference between first and second /third column*
- *XBL condition: difference between first and third column*

Errors

- *No difference between columns in the INC and XBL condition*
- *Difference between first and third column in the XXX and ANI condition (increase)*

Interference

- *XXX as baseline: Difference between first and second column and second and third*
 - *ANI as baseline: No difference between columns*
 - *XBL as baseline: Difference between first and second column*
-

Experiment 2B

Method

Participants

The sample of Experiment 2B consisted of 45 children attending Grades 4-5. Written informed consent was obtained from their parents for their participation.

Material and apparatus

The material and apparatus were the same as in Experiment 2A.

Procedure

The procedure was the same as in Experiment 2A.

Analyses

Analyses were the same as in Experiment 2A, with the exception that for response time, the model formula, in R notation, was specified as:

$$RT \sim \text{column} * \text{condition} * \text{task} + (1 + \text{condition} | \text{subject}) + (0 + \text{task} | \text{subject})$$

The same model was implemented for testing the triple interaction for errors.

Results

Due to software failure two participants were excluded, leaving 43 for analysis.

Results showed that the triple interaction was significant ($\chi^2 = 399.96$, $df = 17$, $p < .001$), indicating that color naming time depended on task, condition, and column. Subsequent interaction analyses with simpler models indicated that the interaction of condition and task was significant ($\chi^2 = 84.30$, $df = 3$, $p < .001$), and so was the interaction between column and task ($\chi^2 = 133.66$, $df = 2$, $p < .000$). The interaction of column and condition was nonsignificant ($\chi^2 = 5.51$, $df = 6$, ns ; Table 1.3).

Table 1.3.

Color naming times and mean number of errors per column in each condition and each task.

Column	Naming Times(s)				Number of errors				
	Card		Single		Card		Single		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
INC									
1	1.12	0.27	0.86	0.15	0.10	0.08	0.11	0.10	
2	1.54	0.41	0.88	0.15	0.13	0.09	0.07	0.07	
3	1.57	0.53	0.89	0.4	0.15	0.10	0.07	0.09	
XXX									
1	0.74	0.18	0.67	0.10	0.02	0.03	0.03	0.03	
2	1.06	0.27	0.72	0.14	0.07	0.05	0.02	0.05	
3	1.07	0.23	0.73	0.13	0.06	0.06	0.02	0.06	
ANI									
1	0.81	0.19	0.76	0.12	0.02	0.03	0.05	0.10	
2	1.22	0.31	0.80	0.12	0.07	0.06	0.03	0.05	
3	1.19	0.26	0.82	0.13	0.05	0.05	0.04	0.04	
XBL									
1	0.85	0.22	0.75	0.11	0.03	0.04	0.03	0.04	
2	1.12	0.28	0.79	0.15	0.06	0.07	0.03	0.05	
3	1.16	0.29	0.80	0.13	0.06	0.06	0.04	0.08	

Note: “INC”: incongruent condition; “XXX”: control condition with repetition of X; “ANI”: control condition with animal names; “XBL”: control condition with repetition of X or B or L.

Post hoc analysis with Bonferroni correction for three comparisons in each condition in each task was employed. In the single trial INC condition, there was no statistically significant difference between columns (all β s > 0.03, $p > .1$). In the single trial XXX condition, there was a statistically significant difference between first and second column, and between first and third (column1 vs. column 2, $\beta = 0.05$, $z = 4.28$, $p < .001$; column1 vs. column3, $\beta = 0.05$, $z = 4.74$, $p < .001$), column2 and column3 did not differ ($\beta = 0.005$, $z = -0.460$, *ns*). The same results were obtained for single trial ANI condition (column1 vs. column 2, $\beta = 0.03$, $z = 2.91$, $p < .05$; column1 vs. column3, $\beta = 0.05$, $z = 4.72$, $p < .001$; column2 vs. column3, $\beta = 0.02$, $z = 1.81$, *ns*),

and for the single trial XBL condition (column1 vs. column 2, $\beta = 0.03$, $z = 2.59$, $p < .05$; column1 vs. column3, $\beta = 0.04$, $z = 3.39$, $p < .001$; column2 vs. column3, $\beta = 0.01$, $z = 0.806$, *ns*; Figure 1.23).

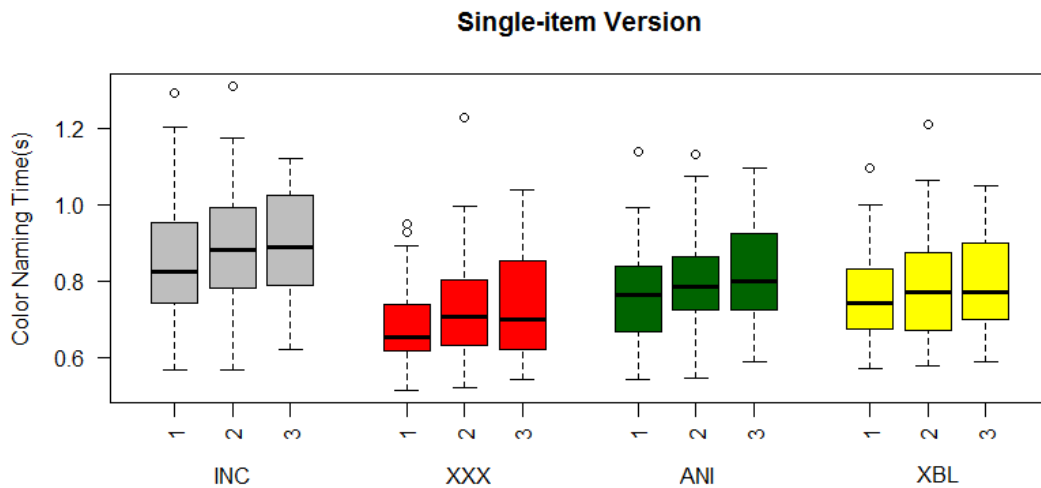


Figure 1.23. Color naming time in the single-item version per column in each condition.

In the card version, for the INC condition, the first column differed from the second ($\beta = 0.41$, $z = 8.18$, $p < .001$), and from the third ($\beta = 0.44$, $z = 8.84$, $p < .001$), but there was no statistically significant difference between the second and third column ($\beta = 0.03$, $z = 0.655$, *ns*). In the XXX and ANI condition, an equivalent pattern was present: the first column differed from the second (XXX: $\beta = 0.31$, $z = 10.03$, $p < .001$; ANI: $\beta = 0.40$, $z = 10.76$, $p < .001$) and from the third (XXX: $\beta = 0.33$, $z = 10.44$, $p < .001$; ANI: $\beta = 0.37$, $z = 9.98$, $p < .001$). Second and third column did not differ in both conditions (XXX: $\beta = 0.01$, $z = 0.412$, *ns*; ANI: $\beta = -0.02$, $z = -0.781$, *ns*). Also, the XBL condition behaved in a similar manner (column1 vs. column 2, $\beta = 0.26$, $z = 10.38$, $p < .001$; column1 vs. column3, $\beta = 0.31$, $z = 12.10$, $p < .001$; column2 vs. column3, $\beta = 0.04$, $z = 1.72$, *ns*; Figure 1.24).

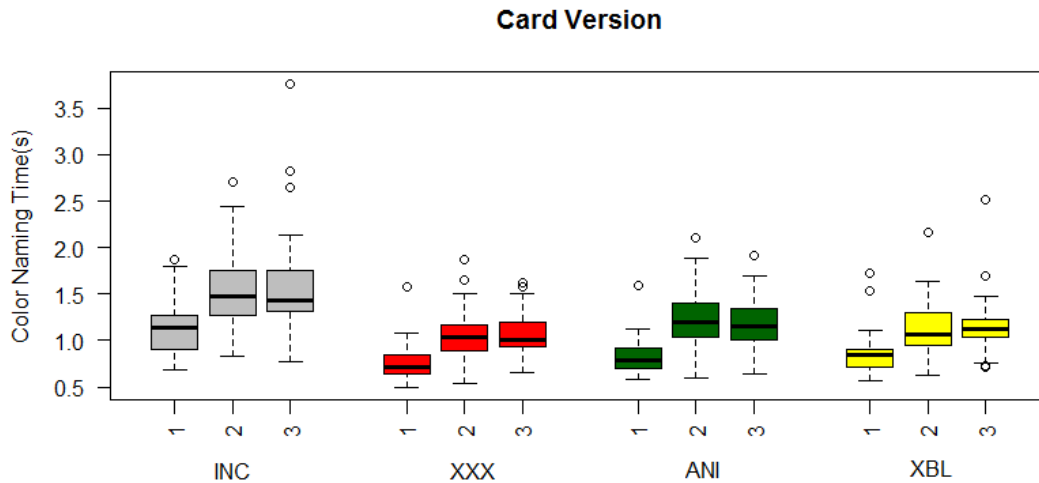


Figure 1.24. Color naming time in the card version per column in each condition.

For interference, the triple interaction $\text{column} \times \text{condition} \times \text{task}$ was significant ($\chi^2 = 26.96$, $df = 12$, $p < .01$). Simpler models revealed that only column and task interacted significantly ($\chi^2 = 17.38$, $df = 2$, $p < .001$).

Separate posthoc pairwise comparisons with Bonferroni correction for interference for each condition and task were performed. For the single trial task, interference for the XXX condition did not differ between columns (column1 vs. column2, $\beta = -0.02$, $z = -1.05$, *ns*; column1 vs. column3, $\beta = -0.02$, $z = -1.05$, *ns*; column2 vs. column3, $\beta = 0.001$, $z = 0.06$, *ns*). The same pattern was observed for the single trial ANI condition (column1 vs. column2, $\beta = -0.008$, $z = -0.449$, *ns*; column1 vs. column3, $\beta = -0.02$, $z = -1.19$, *ns*; column2 vs. column3, $\beta = -0.01$, $z = -0.747$, *ns*), and for the XBL condition (column1 vs. column2, $\beta = -0.006$, $z = -0.322$, *ns*; column1 vs. column3, $\beta = -0.009$, $z = -0.484$, *ns*; column2 vs. column3, $\beta = -0.003$, $z = -0.162$, *ns*; Figures 1.25–1.27).

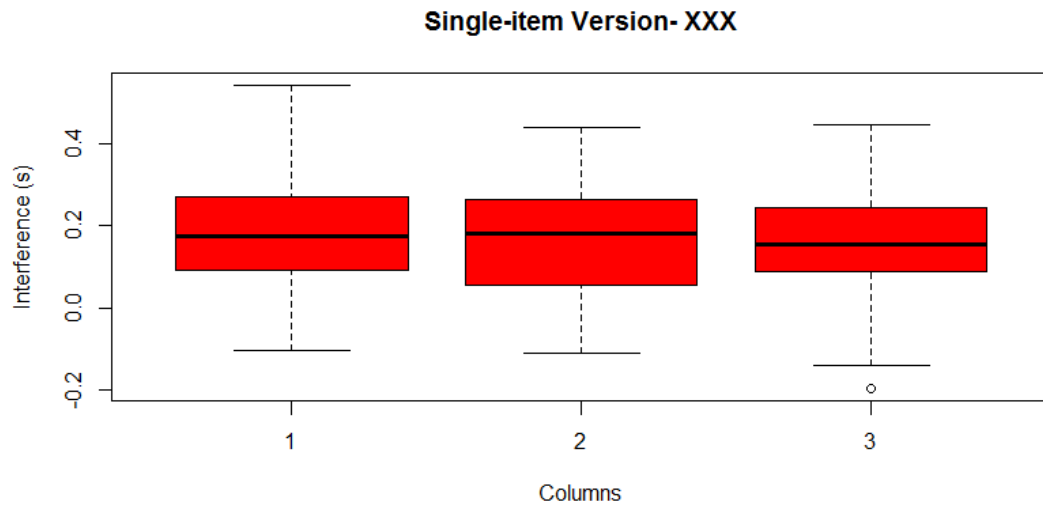


Figure 1.25. Interference in the single-item version using as baseline the XXX condition.

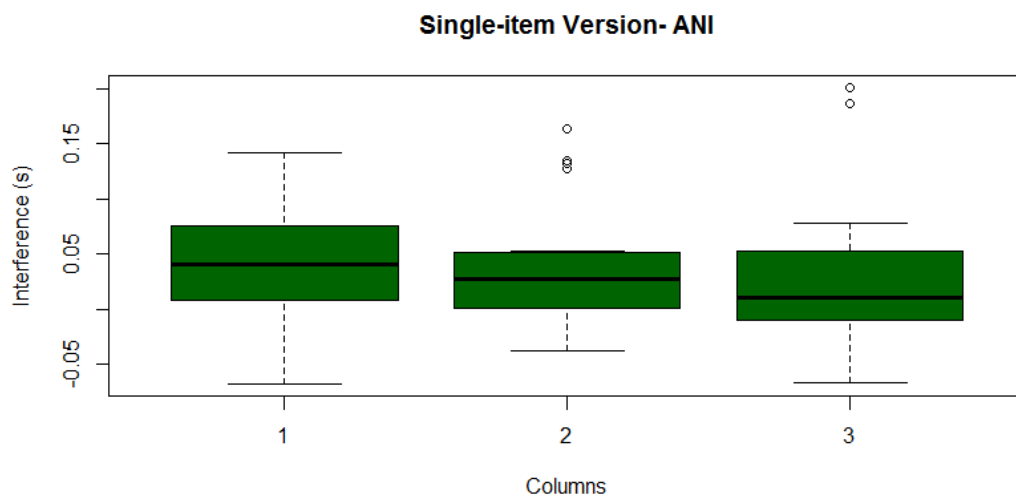


Figure 1.26. Interference in the single-item version using as baseline the ANI condition.

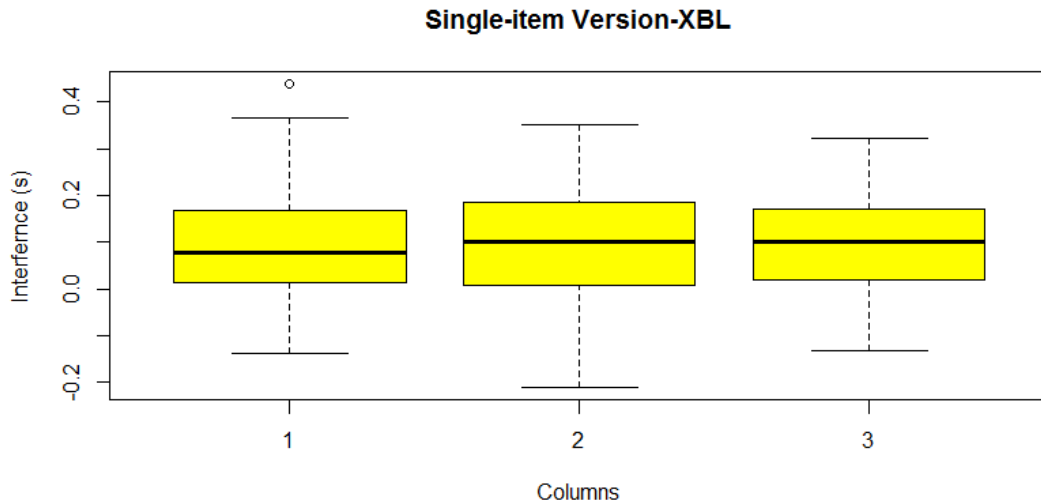


Figure 1.27. Interference in the single-item version using as baseline the XBL condition.

In the card version, for the XXX condition, there was no significant difference in interference between columns (column1 vs. column2, $\beta = 0.09$, $z = 1.80$, *ns*; column1 vs. column3, $\beta = 0.11$, $z = 2.17$, *ns*; column2 vs. column3, $\beta = 0.02$, $z = 0.370$, *ns*). The same pattern of results was obtained for the ANI condition (column1 vs. column2, $\beta = 0.008$, $z = 0.163$, *ns*; column1 vs. column3, $\beta = 0.07$, $z = 1.31$, *ns*; column2 vs. column3, $\beta = 0.06$, $z = 1.15$, *ns*). In contrast, for the XBL condition, interference was statistical different between the first and the second column, and the first and third (column1 vs. column2, $\beta = 0.14$, $z = 2.95$, $p < .001$; column1 vs. column3, $\beta = 0.13$, $z = 2.71$, $p < .05$), but not between the second and third column (column2 vs. column3, $\beta = -0.01$, $z = -0.235$, *ns*; Figures 1.28–1.29).

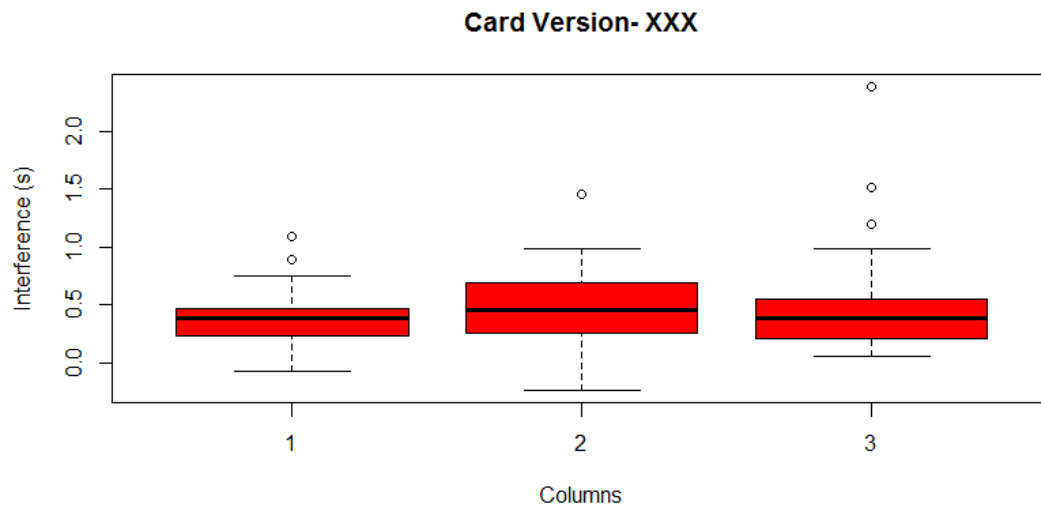


Figure 1.28. Interference in the card version using as baseline the XXX condition.

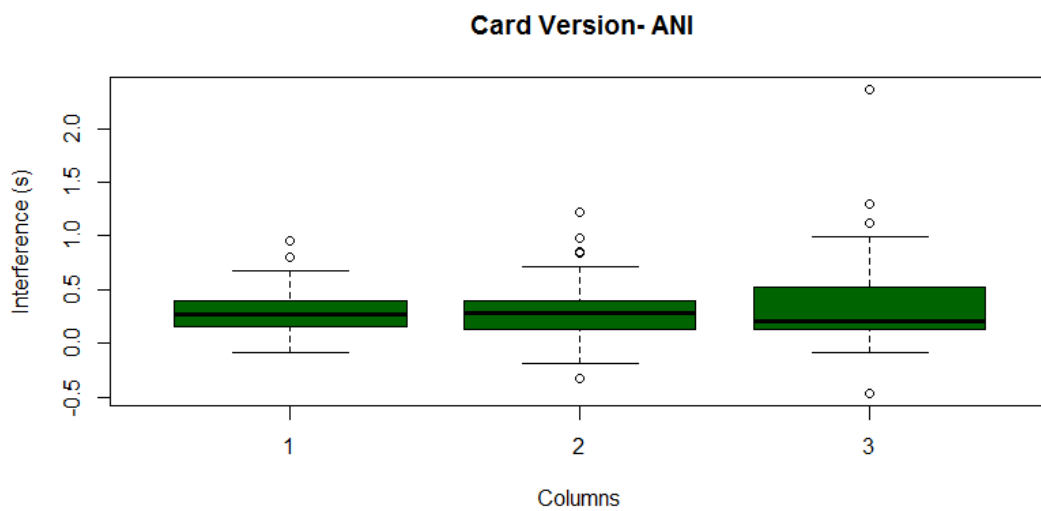


Figure 1.29. Interference in the card version using as baseline the ANI condition.

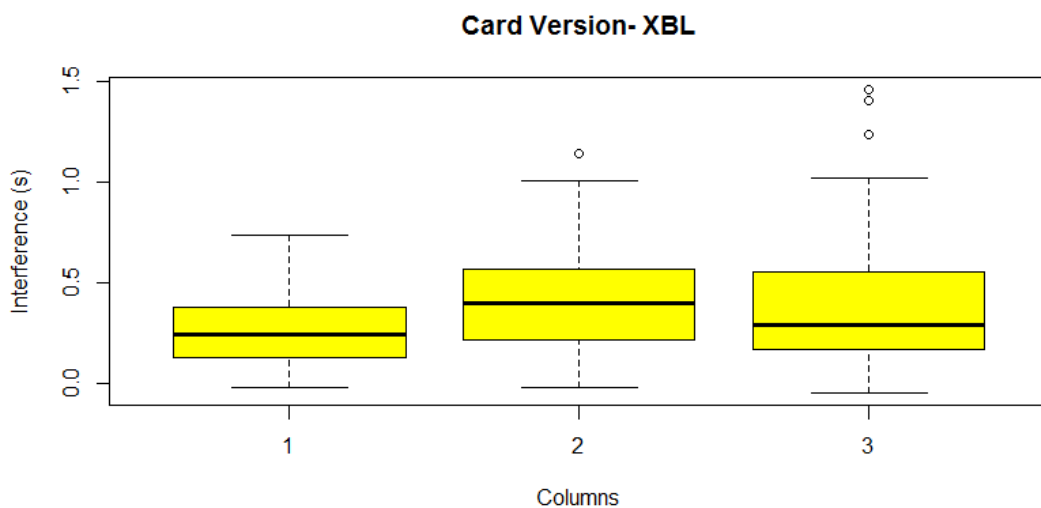


Figure 1.30. Interference in the card version using as baseline the XBL condition.

Since, this time, participants named the control conditions more slowly as both tasks progressed we recalculated the interference for both task versions in each column in each condition, using as the baseline only the first column. The triple interaction $\text{column} \times \text{condition} \times \text{task}$ was significant ($\chi^2 = 210.02$, $df = 12$, $p < .001$). The only interaction that was significant regarded column and task ($\chi^2 = 197.08$, $df = 2$, $p < .001$).

Posthoc pairwise comparisons with Bonferroni correction revealed that for the single trial task, for the XXX condition, there was again no significant difference in interference between columns (column1 vs. column2, $\beta = 0.02$, $z = 1.55$, *ns*; column1 vs. column3, $\beta = 0.03$, $z = 1.97$, *ns*; column2 vs. column3, $\beta = 0.006$, $z = 0.412$, *ns*). Equivalent results were obtained for the ANI condition and the XBL condition (ANI: column1 vs. column2, $\beta = 0.02$, $z = 1.55$, *ns*; column1 vs. column3, $\beta = 0.03$, $z = 1.97$, *ns*; column2 vs. column3, $\beta = 0.006$, $z = 0.412$, *ns*; XBL: column1 vs. column2, $\beta = 0.02$, $z = 1.55$, *ns*; column1 vs. column3, $\beta = 0.03$, $z = 1.97$, *ns*; column2 vs. column3, $\beta = 0.006$, $z = 0.412$, *ns*; Figure 1.31–1.33).

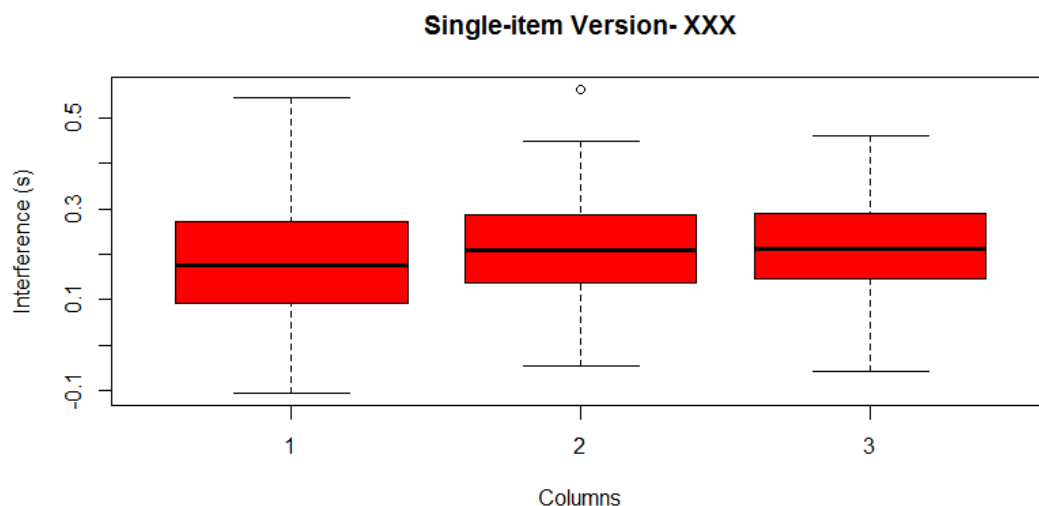


Figure 1.31. Interference in the single-item version keeping XXX baseline constant.

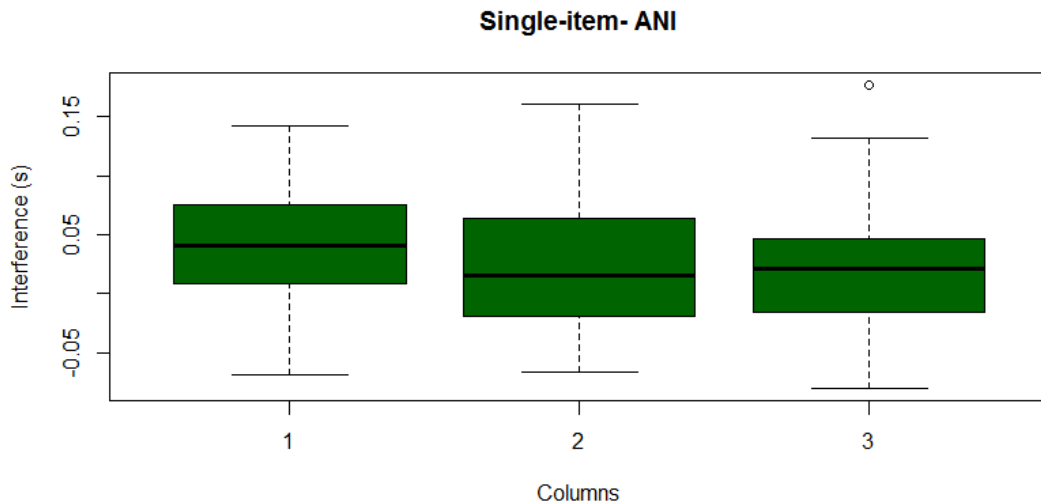


Figure 1.32. Interference in the single-item version keeping ANI baseline constant.

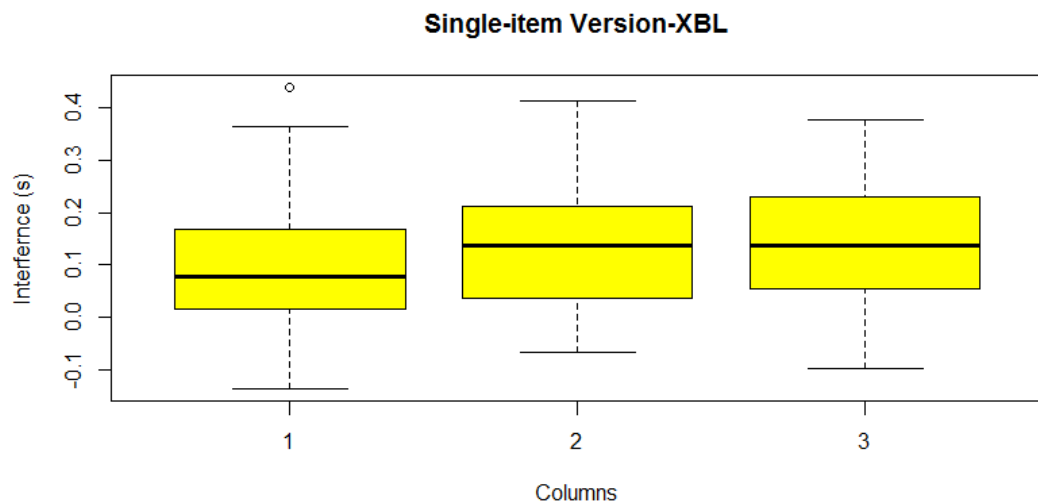


Figure 1.33. Interference in the single-item version keeping XBL baseline constant.

In the card version, for the XXX condition, interference in the first column differed significantly from the second and third (column1 vs. column2, $\beta = 0.41$, $z = 8.18$, $p < .001$; column1 vs. column3, $\beta = 0.44$, $z = 8.84$, $p < .001$). No difference was observed between the second and third column (column2 vs. column3, $\beta = 0.03$, $z = 0.655$, *ns*). The same results were found regarding the ANI condition (column1 vs. column2, $\beta = 0.41$, $z = 8.18$, $p < .001$; column1 vs. column3, $\beta = 0.44$, $z = 8.84$, $p < .000$; column2 vs. column3, $\beta = 0.03$, $z = 0.655$, *ns*), and the XBL condition

(column1 vs. column2, $\beta = 0.41$, $z = 8.18$, $p < .001$; column1 vs. column3, $\beta = 0.44$, $z = 8.84$, $p < .001$; column2 vs. column3, $\beta = 0.03$, $z = 0.655$, *ns*; Figures 1.34–1.36) .

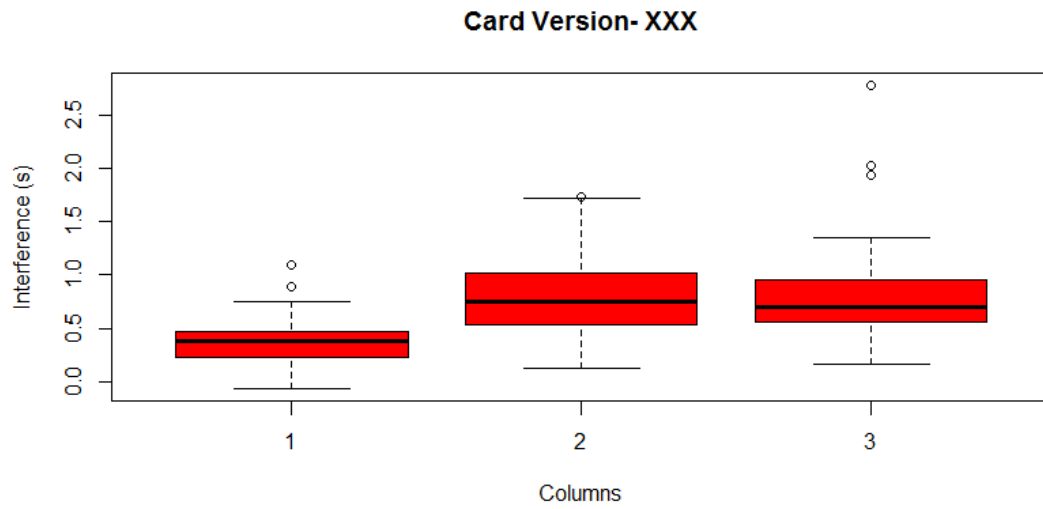


Figure 1.34. Interference in the card version keeping XXX baseline constant.

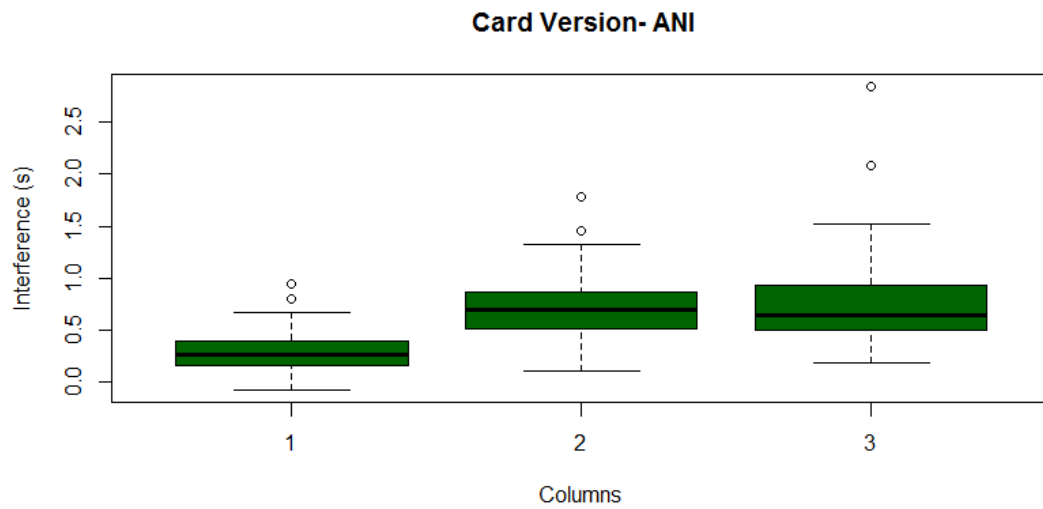


Figure 1.35. Interference in the card version keeping ANI baseline constant.

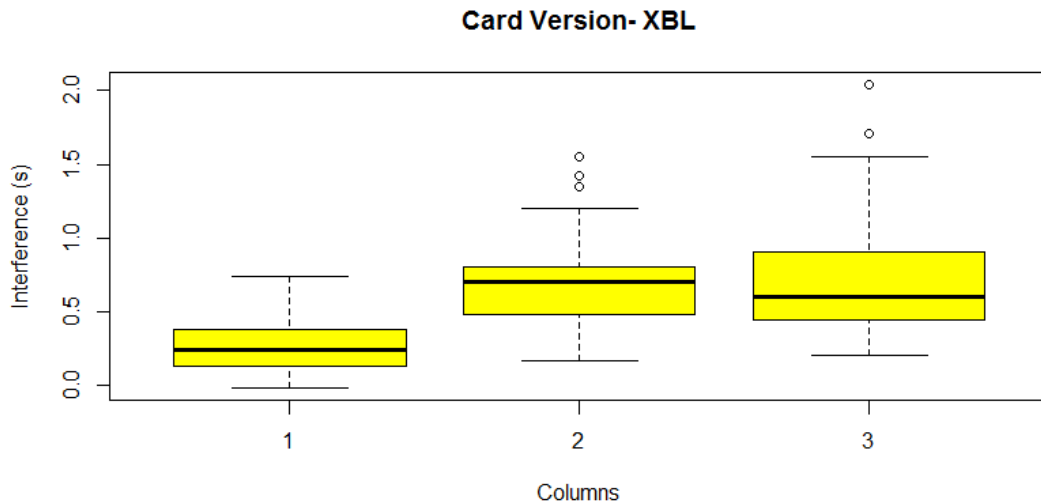


Figure 1.36. Interference in the card version keeping XBL baseline constant.

For accuracy data (mean number of errors) we employed a general linear mixed-effects model. The results showed that the triple interaction of column \times condition \times task was significant ($\chi^2 = 70.03$, $df = 17$, $p < .001$). Subsequent analysis showed that condition and task interaction was significant ($\chi^2 = 17.41$, $df = 3$, $p < .001$). The same was observed for column and task ($\chi^2 = 38.70$, $df = 2$, $p < .001$), but not for column and condition ($\chi^2 = 6.56$, $df = 6$, ns).

Post hoc analysis with Bonferroni correction for three comparisons in each condition in each task for each column was employed. For the single trial task, in the incongruent condition, errors dropped from the first to the second and third column. But not between second and third (column1 vs. column2, $\beta = -0.04$, $z = -2.78$, $p < 0.05$; column1 vs. column3, $\beta = -0.03$, $z = -2.28$, ns ; column2 vs. column3, $\beta = 0.007$, $z = 0.500$, ns). For the XXX condition, there was no difference in errors between columns (column1 vs. column2, $\beta = -0.005$, $z = -0.858$, ns ; column1 vs. column3, $\beta = -0.009$, $z = -1.43$, ns ; column2 vs. column3, $\beta = -0.003$, $z = -0.575$, ns). The same pattern was observed in the ANI and XBL conditions (ANI: column1 vs. column2, $\beta = -0.02$, $z = -1.73$, ns ; column1 vs. column3, $\beta = -0.01$, $z = -0.90$, ns ; column2 vs. column3, $\beta = 0.01$, $z = 0.$, ns ; XBL: column1 vs. column2, $\beta = 0.002$,

$z = 0.25$, *ns*; column1 vs. column3, $\beta = 0.009$, $z = 1.02$, *ns*; column2 vs. column3, $\beta = 0.007$, $z = 0.774$, *ns*; Figure 1.37).

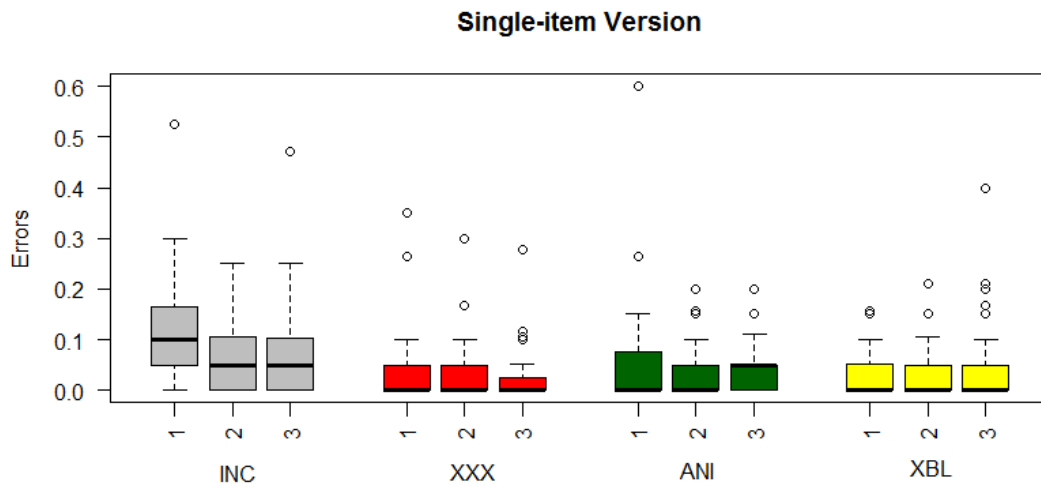


Figure 1.37. Mean number of errors in the single-item version per column in each condition.

In the card version, for the INC condition, errors were higher in the second and third column compared to the first (column1 vs. column2, $\beta=0.03$, $z = 2.60$, $p <.05$; column1 vs. column3, $\beta= 0.04$, $z =3.53$, $p <.01$), but the second and third column did not differ (column2 vs. column3, $\beta= 0.01$, $z = 0.931$, *ns*). The same pattern of results was observed for the XXX condition (column1 vs. column2, $\beta = 0.05$, $z = 4.78$, $p <.001$; column1 vs. column3, $\beta = 0.03$, $z = 3.69$, $p <.001$; column2 vs. column3, $\beta = -0.01$, $z = -1.08$, *ns*), for the ANI condition (column1 vs. column2, $\beta = 0.04$, $z = 4.59$, $p <.001$; column1 vs. column3, $\beta = 0.02$, $z = 2.46$, $p <.05$; column2 vs. column3, $\beta = -0.02$, $z = -2.13$, *ns*), and for the XBL condition (column1 vs. column2, $\beta = 0.03$, $z = 2.59$, $p <.05$; column1 vs. column3, $\beta = 0.03$, $z = 2.79$, $p <.05$; column2 vs. column3, $\beta = 0.002$, $z = 0.20$, *ns*; Figure 1.38).

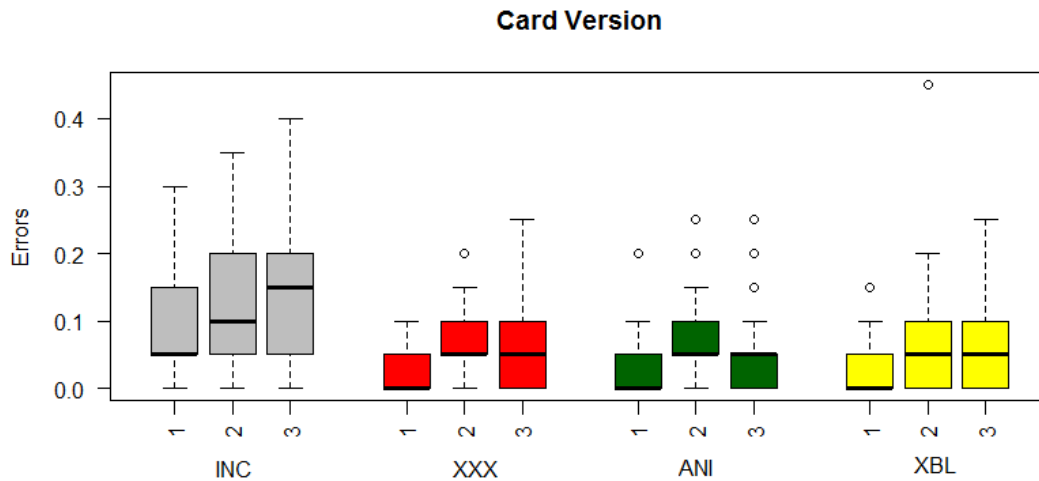


Figure 1.38. Mean number of errors in the card version per column in each condition.

For children, the substitution errors for the XXX condition were 37.5% of the total errors, for the ANI condition 27.81%, and for the XBL condition 28.78%. For the ANI condition, the reading errors were 8.28%.

For summary of the main findings in Experiment 2B, see Box 1.3.

Discussion

For adults the same pattern of results was obtained as in Experiment 1 for the incongruent condition. There was a steep increase in color naming time between the first and second, and third columns in the incongruent condition. Also, a gradually decline in color naming time as the task progressed, was observed for all three control conditions.

The results for children were generally similar, although somewhat different for the control conditions. Decrease in color naming time was equally steep in all four conditions, control conditions being no exception.

The results suggest that the influence of multiple sources of interference, and that the effects of stimulus similarity, and physical identity on performance may differ between developmental stages.

Box 1.3

Main findings in each condition and task.

Single-item version

Color naming

- *No difference between columns in INC condition*
- *Difference between first and second /third column in all control conditions*

Errors

- *Difference between first/second column in INC condition (decrease)*
- *No difference between columns in all control conditions*

Interference

- *No difference between columns in all conditions*

Card version

Color Naming

- *Difference between first and second / third column in all conditions*

Errors

- *Difference between first and second / third column in all conditions (increase)*

Interference

- *XXX as baseline: No difference between columns*
 - *ANI as baseline: No difference between columns*
 - *XBL as baseline: Difference between first and second / third column*
-

General Discussion

The aim of the present study was to examine different hypotheses deriving from the conflict monitoring model and the strength model of self-control. According to the first theory, adjustments in control should result in a performance improvement during the Stroop task in the incongruent condition, since it is proposed that successive conflicting stimuli raise the alertness of the control system, leading it to adjust to the requirements of the task. In contrast, the strength model of self-control

suggests that, due to the limited nature of self-control, decrements in performance emerge in conflicting tasks, and the individual enters an ego-depletion state, where self-control can no longer be exerted. Our findings contradict the former theory, but they support and, in my view, expand the latter theory.

The steep increase in color naming time for the incongruent condition observed in all three Experiments, and in two different developmental stages, suggests that the engagement in an extremely brief, but effortful and complex task suffices to produce ego depletion effects. Depletion was evident within seconds, and this finding implies that prolonged task duration is not required to produce these effects, as task complexity suffices.

We also aimed to test two further assumptions of the self-control model, namely the recovery and conservation hypotheses. Our findings are consistent with the recovery hypothesis, suggesting that self-control can be impaired, when a total lack of rest periods blocks replenishment. Rest times thus seem crucial for self-control, in contrast to task duration, which seems unimportant. Previous failures to confirm the recovery hypothesis (Hagger et al., 2010) may have been, due to the dual-task paradigms used, and the way tasks are implemented in self-control research (e.g., simple-item instead of card version of Stroop task). Specifically, the unintended and uncontrolled interim periods between stimuli and between tasks may have diminished ego-depletion effects.

The notion of conservation assumes that expectation of subsequent self-control acts causes conservation of resources (Baumeister et al., 2007). In our task, the second column should thus be affected more than the third column, as participants should curtail effort more aggressively to be able to exert self-control in the third column. This was not observed. In contrast, our findings are consistent with a limit to ego

depletion beyond which self-control is no longer effective (Baumeister et al., 2007), leading to impaired yet stable self-control.

One of the most interesting findings of our study was that color naming time increased in the control condition as well. This finding cannot be explained by either theory, and especially not by the conflict monitoring hypothesis. We have proposed that in card version of the Stroop task, there are multiple sources of interference, deriving from adjacent stimuli, and we have argued that, if this were the case, then it would also be evident in the control condition. The results of the present study are consistent with this hypothesis. Unless we reconsider the way the control condition is processed, we cannot explain the observed pattern of results. Thus, a reconsideration of the proposed model of Botvinick et al (2001) seems in order.

Moreover, in my view, ego-depletion theory and the conflict monitoring hypothesis can be combined in light of the above evidence, and others previously reported. Specifically, Inzlicht and Gutsell (2007) examined the neural basis of self-control, and in an EEG study, they found that Error Related Negativity, a waveform associated with the activity of ACC, was attenuated when participants were depleted compared to non-depleted participants. This suggests that ACC activation can be weakened after depletion, and the performance impairment can be associated with this diminished activation.

Therefore, we propose a modification of the model of Botvinick et al. (2001) to account for the data observed in our study regarding the Stroop task (Figures 1.39 and 1.40). In this model, a layer is added that corresponds to the spatial location of the stimuli. Input layers take activation not only regarding the color and word dimensions of a stimulus, but also regarding their spatial location. If a target “red” is to be named, then it activates the units that correspond not only to the color and word, but also the spatial location of the stimulus. The conflict monitoring unit is connected to the task

demand and spatial location layer, and defines the adjustments of control. Also, a limit is defined for the conflict monitoring unit, after which activation drops, and control is impaired. Activation in the conflict monitoring unit at time n is added to activation of this unit at time $n+1$, so continuous effort for control increases in subsequent trials, and the activation of the conflict monitoring unit begins to drop rapidly and abruptly. For the control condition, where nearby stimuli are the sole source of interference, conflict appears only regarding the spatial location of the stimulus. Consequently, the activation added in the conflict monitoring unit is more attenuated, and performance drops more gradually, at least in the case of adults. For children, it is possible that one source of interference is sufficient to produce steep performance decrement. Overall, our findings suggest that the rate of dropping may depend on task complexity, individual differences, and maturity of the conflict monitoring system.

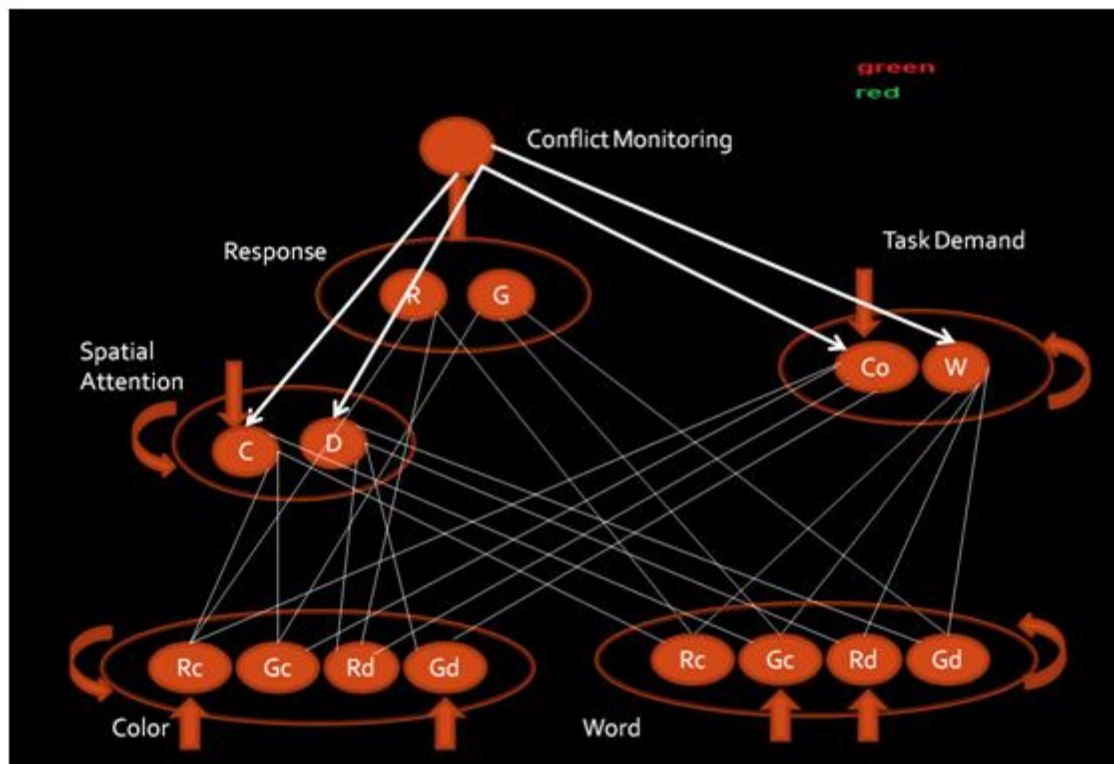


Figure 1.39. Simulation of card version of Stroop task for two adjacent color and word responses (incongruent condition).

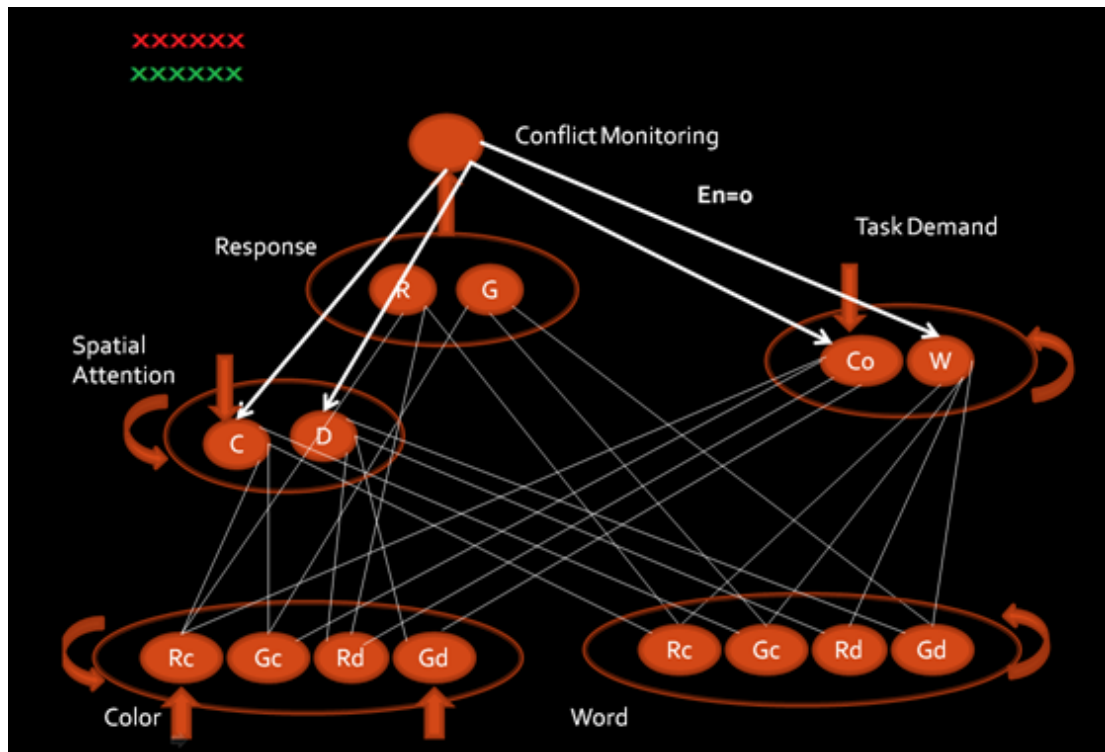


Figure 1.40. Simulation of card version of Stroop task for two adjacent color responses (control condition).

The model proposed above needs further investigation with fMRI studies and eye-tracking techniques. First, as we have presented in the introduction, the vast majority of the studies implementing fMRI have used the single-trial version of the Stroop task. ACC activation decrement, during the card version of the Stroop task, has to be established in the future. In addition, the processing of adjacent stimuli in the card version of the Stroop task, can be enlightened by eye-tracking investigation of this version. Unfortunately, eye-tracking studies of the Stroop task have examined the single item version and focused on alterations of pupil size during conflict (Laeng, Ørbo, Holmlund, & Miozzo, 2011) and first fixation factors (Perret & Ducrot, 2010), and not eye-movements, gaze duration, fixation duration, eye-voice, and perceptual span, etc.

One objection that can be raised against our interpretation, is that children manifested slower naming times in the single trial version of the Stroop task in the control conditions as the task progressed, and, consequently, it is the number of items

that produce the pattern observed. This would be plausible, if it was also observed in the incongruent condition, but it was not. It should be noted, that the duration of all single trial tasks was 12min, so it is reasonable to assume that children got tired to be involved for such a long time in a non-challenging task like that of the control conditions. Performance in the incongruent condition, which poses more demands to the children, enhances this possibility.

In the present study, we additionally showed how the course of the control condition may affect the observed interference. This is a critical point, because it outlines how the baseline chosen, can distort the estimated interference, and suggests that subtraction, as a method to obtain interference, may be inadequate, and even misleading. Although this is in line with the proposal of Salo et al. (2001), our findings run counter to the other main conclusions of their study, and to the expectation that in the card version color naming is facilitated because participants can look ahead, and thereby speed up their responses (MacLeod, 2000). Specifically, the card version of the task was not faster than the single trial version in any of the three Experiments; in fact the opposite was true. More importantly, it was not the speeding in the responses within the control condition of the card version that produced the magnitude of interference and its difference between task versions. Our data suggest that the difference in interference observed between task versions is, due to the steep increase of color naming times as the card version progresses in the incongruent condition, which does not occur in the single trial version. Two observations are of interest: First, in Experiment 2A, involving adults, the mean response time was similar in both task versions in the first column for all conditions. And second, in both Experiments, the second and third column of the control condition of the card versions reached the response times of the incongruent condition

of the single trial version. Although the last comparison may be not straightforward, it is worth noting, and acknowledged in future studies.

To the extent that the card version of the task is usually implemented in clinical settings as a marker of inhibition and selective attention deficits, it is crucial to determine how depletion and variations of performance within the incongruent, but also within the control condition, define the conclusions drawn for a variety of disorders. For example, it is well established that people with schizophrenia tend to present an abnormal interference, when measured with the card version of the Stroop task, and normal interference, when measured with the single trial version (Henik & Salo, 2004). If we do not establish, how depletion affects different kinds of populations, and how it impacts them in each condition (i.e., in incongruent vs. control condition), and, ultimately, what exactly we measure, when applying the Stroop task, then all statements are moot, and unwarranted conclusions are drawn. In addition, this issue challenges the comparability of our research: how can studies that use different variations of the Stroop task—not only regarding the versions but also the control conditions—be meaningfully compared? If we are not confident in what we measure, we cannot be confident in our comparisons. Future research should clarify these issues.

An additional goal of our study was to test possible crowding effects in the control condition as an alternative for the results of Experiment 1. What we observed was that in both populations the XXX control condition was faster than the other two conditions, with the XBL conditions being the slowest. A hasty interpretation would be that crowding had no effect on participants' performance. Instead, a closer look into our findings reveals that although the XXX condition was the fastest, it exhibited greater increase in color naming times in adults than the XBL condition, where stimuli were form (shape) dissimilar. This suggests that although crowding and form

similarity may have no effect on the initial stages of the task (i.e., column1), and may even be facilitatory, it is an additional source of depletion, and causes a steeper decline in performance. Although the adult sample was small, and no final conclusion can be reached, this pattern was not evident in children, in line with observations that crowding is more pronounced in children (<11 years old), and, consequently, had a steep increase in color naming in all control conditions (Jeod, Hamid, Maurer, & Lewis, 2010) . To the best of my knowledge, there is no study examining the continuous performance in crowding tasks in order to study its course. Future studies should establish this issue in more depth.

A few minor points regarding our study should be addressed. First, one could claim that an alternative to the limited capacity of self-control, as manifested in the last two columns of the card version of the Stroop task, is that participants adapted to the situation, and found a balance. If this were the case, then we should have observed no difference between the second and the third column in the control condition of Experiment 1. Also, the XBL condition should not have produced an increasing trend, as it did, at least for adults in Experiment 2, although we acknowledge that a larger sample is needed in order to be conclusive.

Second, mental fatigue that is a state of sub-optimal functioning, the experience of tiredness and the loss of commitment to the task after or during the engagement in a demanding cognitive task (Boksem & Tops, 2008), can be excluded as an differential explanation. In order for mental fatigue to be induced a prolonged duration of cognitive activity is in principle required (Boksem & Tops, 2008). Since the card version of the Stroop task is extremely brief, and duration is a precondition for mental fatigue, such an explanation cannot be applied in our study to account for the pattern observed. Moreover, even in the incongruent condition of the single trial version with significantly longer duration, no such pattern was observed. If this theory

stands as an explanation for the control conditions of the single trial task, then, probably, it needs redefinition since this task is not thought to be demanding, and seems not to be.

Study 2: Pause time

RAN-Color and Stoop Task: Twins, siblings, or distant relatives?

In 1960 a neurologist named Norman Geschwind described the case of a patient, who demonstrated pure alexia without agraphia. Geschwind spoke for a visual–verbal disconnection, and went further with the investigation of a different patient, who, although capable to perceive colors, was unable to name them. For his examination Geschwind developed a timed color naming task, which consisted of an array of 50 colored squares in five rows in five different familiar colors. In 1972, a student of Geschwind, M. Denckla, was interested in exploring the relationship between color naming and reading, and studied a large group of kindergarteners. What she found was that five boys with dyslexia produced prolonged color naming times, which suggested, that continuous color naming and reading may have a common ground. Denckla, together with Rita Rudel, created additional variants of the initial color naming task using alphanumeric stimuli (i.e., letters and numbers) and another class of nonalphanumeric stimuli, namely objects. They termed these tasks “Rapid Automatized Naming” (RAN) for describing the speeded naming of highly familiar objects (Denckla & Cutting, 1999; Norton & Wolf, 2012).

Since then many studies have implemented RAN in the study of reading development and reading disorders. The format usually adopted is that of 50 items of familiar objects arranged in five rows of ten items, presented pseudorandomly, with no same items appearing successive (Figures 2.1 and 2.2). Participants are required to name the stimuli as quickly as possible in a left-to-right fashion (Norton & Wolf, 2012).

a o s d p a o s p d
o s a p d a s d o p
s p o d a p o a s d
d a p o s a s d p o
p d a s o p d o a s

5 3 1 7 9 3 5 9 7 1
1 7 3 9 5 7 9 1 5 3
3 5 9 1 7 5 3 9 1 7
7 1 5 3 9 1 5 3 7 9
9 5 1 7 3 9 1 7 3 5

Figure 2.1. Alphanumeric RAN-test.

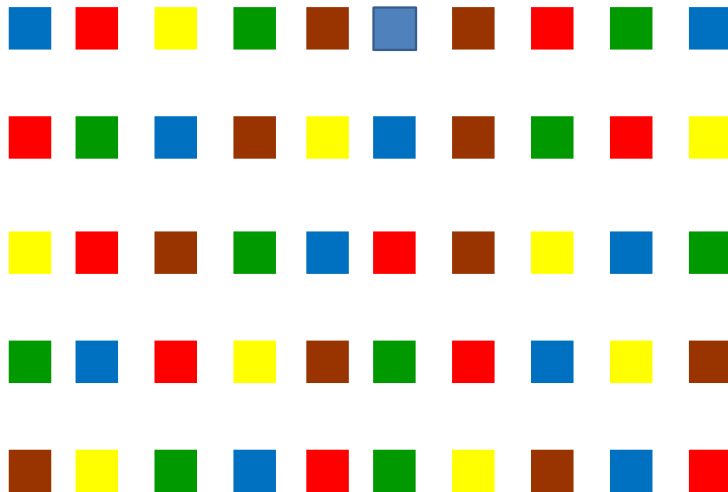


Figure 2.2. Nonalphanumeric RAN-test.

It is well established that RAN predicts reading acquisition, and can distinguish between fluent and poor readers. For example, Kirby, Parrila, and Pfeiffer (2003) in a longitudinal study found that naming speed, as measured with the RAN task, contributes independently from other factors, such as phonological awareness, to the prediction of reading, and in fact is a more powerful predictor, when examining latter stages of reading, after Grade 1. The same pattern had been also observed earlier by Scarborough (1998), who examined children in Grades 2, and made follow up

measurements in Grade 8. In this research, RAN was found to predict reading development, especially in children with reading disabilities.

More important for the present study is the finding that RAN is also a predictor of reading level in adulthood as well. Felton, Naylor, and Wood (1990), using all four variants of RAN (letters, numbers, colors, and objects), found that adults with a history of reading disability performed more poorly in rapid naming than skilled readers, and this task was a powerful discriminator between groups. Similar results were reported by Korohnen (1995), who examined children 9 years old with reading disabilities, and followed them to age 18. They administered the Rapid Alternating Stimulus Naming (RAS)— an equivalent task to RAN, where two or three items of different classes are repeated alternately (e.g., in the Korohnen study one RAS task included letters and numbers, and another included letters, numbers and colors). The results showed that impairments in RAN performance persisted into adulthood, and the authors argued in favor of a domain-general deficit mirrored in naming performance, which includes motor, phonological, and orthographic components.

Despite the above claim, it is broadly accepted in the reading literature that, although RAN is strongly related with reading ability, it is not evident why this relationship exists. One of the problems is the multicomponential nature of RAN. As Wolf & Bowers (1999) propose, there are seven main components that are reflected in RAN performance: a) attention, b) visual processing, c) integration processes of visual features with orthographic representations, d) integration of visual features with phonological representations, e) phonological and semantic access, f) integration and activation of that information, and g) motoric processes for articulation. Which of all the above components are reflected in the relationship of RAN to reading, remains unresolved.

Furthermore, an additional issue that may distort our observations and conclusions drawn, is the way naming speed is measured in the RAN tasks. Specifically, most researchers measure the total naming time of the task. This method, although widely adopted, may be inadequate to enlighten the influence of RAN on reading, because it cannot illuminate specific sub-processes such as attention, shifting, articulation because of its serial nature (Georgiou, Parrila, & Kirby, 2006). In addition, although naming speed and RAN are used interchangeably in the literature, this is misleading, because speed expresses the number of correct units per time, where RAN is not always rapid, since errors are also included (Kirby, Georgiou, Martinussen, & Parrila, 2010). The inclusion of articulation of errors in the total naming time is an inherent problem of all the serial tasks, and, consequently, of RAN, and may affect the observations and inferences made.

In light of the aforementioned considerations, researchers aimed to establish, which intratask components may contribute to the relationship of RAN with reading, and broke the task based on articulations and pause times. The articulation time is the amount of time needed to articulate each stimulus, and pause time refers to the silent gaps between stimuli (Norton & Wolf, 2012). Although this issue is fundamental for the reading literature, the studies examining the impact of both components on reading ability are limited.

The first evidence for the relationship of RAN with the above intratask components derive from the study of Anderson, Podwall, and Jaffe (1984). In their study six 8-to-10 year-old children with dyslexia were included, and were tested in all RAN tasks, specifically with colors, objects, and letters. Articulation and pause times were estimated, and they found that children with dyslexia with reference to a control group exhibited prolonged time duration in both components. They concluded that

children with dyslexia had greater total naming times in the RAN tasks not only because they paused more, but also, because they articulated more slowly.

In contrast, in the study of Obregon (as cited in Georgiou et al., 2006) a different pattern was observed. He administered RAN using letter, objects, and colors in readers with dyslexia and average readers, and calculated additional variables along with articulation and pause time, i.e., time spent on errors and end-of-line scanning. His results indicated that only pause time differed significantly between groups.

Cobbold, Passenger, and Terell (2003) in a longitudinal study examined the variability in an object RAN task in children across different developmental stages. At the beginning of the study the age of the children ranged from 48 to 54 months, and the study lasted 12 months. Children were tested twice, and measurement of word and nonword reading was also included in the last session. The results suggested that, during the initial assessment, children manifested variability regarding the naming times in the RAN task that seemed to be induced by the different pause times and not articulation times. Also, in the second administration of the RAN task, pauses times and not articulation times were related to the reading tasks.

More recently, Georgiou et al. (2006) examined articulation and pause time in children before and after reading acquisition (kindergarteners to Grade 1). In this study, RAN-C and RAN-N, as also reading fluency and reading accuracy measures, were included. Their findings suggested that articulation time did not develop between measurements, and was less stable. In contrast, pause time was not only stable, but developed significantly, and correlated more strongly with reading measurements.

After indicating that pause time is a stronger predictor for reading than articulation time and other measures, the critical question for the researchers is what pause time reflects. Two main alternatives have been proposed.

Wolf and Bowers (1999) proposed a new conceptualization of developmental dyslexia regarding among others evidence related to the RAN tasks. Although their main goal was not to approach a theory on pause times, they expressed an opinion that still remains as an alternative for pause times. Specifically, referring to the aforementioned study of Obregon (1994), they suggested that pause time may reflect the additional time needed by participants with dyslexia to abandon processing of the previous item, and engage to the processing of next one. In other words, pause time reflects inhibition, in the sense, that during the silent gaps between items participants inhibit the connection of an item and its name in order to move to the next item.

In contrast, Neuhaus, Foorman, Francis, and Carlson (2001) proposed that pause time is related to retrieval. In their study they investigated the relationship of pause and articulation time with reading. They administered three RAN tasks (letters, numbers, and objects) and reading measurements in 50 children, attending first and second grade, in two different time points. Their results indicated that pause time for the RAN-letters task, in contrast to RAN-numbers and RAN-objects, was a stable predictor of reading. They proposed that pause time for letters reflects a specific verbal speed of processing associated with letters, where pause time in other tasks, such as object naming, a more general verbal processing speed. In their view, retrieval of the information needed to complete the task is evident in pause times. More important for the purposes of the present study is the notion that, whereas in children less stable memory representations and retrieval processes coexist, for adults it is only retrieval processes that are evident in RAN. This is in line with studies treating RAN as a retrieval measure (Wolf, Bally, & Morris, 1986).

For nonalphanumeric RAN tasks, it is known that color naming is a better predictor for reading than object naming. For example, Blachman (1984) found that color naming was strongly correlated with reading measurements, whereas object

naming was only moderately so. This evidence is also supported by a longitudinal study (Wolf et al., 1986). Georgiou et al. (2006) examined pause and articulation time in RAN color, and found that it was related with a variety of reading measures in different time points. Regarding these results, it seems justified to hypothesize that processes of inhibition or retrieval can be implemented in RAN-color as approaches for pause time.

It is generally thought that the Stroop task resembles the RAN tasks, and shares characteristics with them (Norton & Wolf, 2012). Also, the relationship of word reading ability and the Stroop task is well established. Everatt et al. (1997) found that children with dyslexia exhibit more interference than age-matched controls. More recently, Protopapas, Archonti, and Skaloumbakas (2007) showed that reading ability is negatively related to Stroop interference. In a first study, where children with dyslexia were compared to age-matched controls in the Stroop task, the results showed greater interference for the children with dyslexia. In their second study, examining the relationship between interference and reading skills in the general school population, they found that poorer readers produced more interference. Furthermore, interference was found to be primarily associated with reading speed. Protopapas et al. suggested that reading ability and interference are directly linked.

In addition, in one study, where RAN, reading, and executive functions were investigated the control condition of the Stroop task was treated as an equivalent of the RAN-color task (Stringer, Toplack, & Stanovich, 2004). In this study, the card version of the task consisted of 48 color stimuli in four colors presented in a matrix of six columns of eight rows. The results showed that RAN-letter and RAN-numbers were correlated with reading, but RAN-colors (i.e., the control condition of the Stroop task) did not correlate with reading, but only with executive function measures. Even though the above imply that color naming in Stroop and in RAN have same features,

and involve similar processes, to the best of my knowledge no study has directly compared the two color naming tasks to examine their relationship.

Aim of the Present Study

Our goal was to investigate, and test both alternative explanations regarding pause time, namely inhibition vs. retrieval. Since the Stroop task is considered equivalent to RAN tasks, and we were primarily interested in pause time as reflected in the Stroop task, Stroop stimuli were used.

Our main manipulation involved the presentation duration of stimuli in the single trial version of the Stroop task. In one task, stimuli were removed from the screen immediately after the participants have completed their articulation of the response, presumably leaving no time for inhibition (SPEEDED-T), whereas in the second task, stimuli were presented for a fixed-time duration, giving them the opportunity to inhibit their response (SLOW-T).

If pause time reflects “the time taken to inhibit the previous symbol and its name and then processing the visual and lexical information about the new symbol” (p. 419), as Wolf & Bowers (1999) propose, then we should expect that response times in SPEEDED-T will be slower compared to SLOW-T, because participants will need extra time to inhibit the previous stimulus, and process the next. In addition, inhibition measurements should correlate more strongly with the response times of SPEEDED-T in both conditions (incongruent and control), and the same should be true for word reading and RAN measurements.

If, on the other hand, retrieval (i.e., verbal processing speed), is reflected in pause time, as Neuhaus et al. (2001) suggest, then no difference in response times should be evident between the two task versions, since the retrieval time will be independent of presentation duration. Regarding inhibition and word reading, both versions of the task should correlate equally with those measurements.

A secondary and supplementary goal of our study was to test the assumption that the Stroop task and RAN-color involve common processes. Although this speculation seems plausible, due to the nature of the tasks, which are based in speeded color naming, and as we have demonstrated with the study of Stringer et al. (2004), can be used interchangeably, to the best of my knowledge, no study has directly examined their relationship, and, ultimately, their relationship to reading.

Method

Participants

The total sample consisted of 41 adults 21-38 (32 female), mainly undergraduate and graduate students. All participants were native speakers of Greek.

Material and Apparatus

Naming measures

Stroop Task. The Greek words for red (κόκκινο /kocino/), green (πράσινο /prasino/), and yellow (κίτρινο /citrino/) were used, because they have the same number of letters and syllables, comparable written frequency, and begin with voiceless stops, which facilitate response time triggering. The corresponding colors are familiar and easily distinguishable. Stimuli for the neutral color condition were made up of 7 repetitions of the letter X (no spaces) in red, green, and yellow color (RGB #FF0000, #00FF00, and #FFFF00, respectively). For the incongruent condition, the Greek words for red, green, and yellow appeared in a nonmatching color.

Each task consisted of 24 stimuli, which were presented randomly on a 15.5'' laptop screen in black background 40-pt Arial font. Practice trials preceded data collection.

For the speeded version (SPEEDED-T) of the Stroop task, the stimuli appeared on the screen, until the participants articulated their response (Figure 2.3). As soon as

the participants completed their response, the experimenter, who was sitting next to them, clicked on the mouse, and the next stimulus was presented on the screen. SPEEDED-T lasted approximately 45sec for the control condition ($M = 45.08$, $SD = 2.29$), and 49sec for the incongruent condition ($M = 49.11$, $SD = 3.37$).

For the version with slow administration time (SLOW-T), the stimuli stayed on the screen for 2sec without any intervention of the experimenter, and were removed automatically (Figure 2.4). The duration of 2sec suffices to produce a response and a temporal gap to intervene between trials. SLOW-T lasted 64sec in both conditions.

Due to software constraints, there was an interstimulus interval of 166.67ms in both tasks.

The stimuli of the card version of the Stroop task, as presented in Experiment 1 of Study 1, were also used.

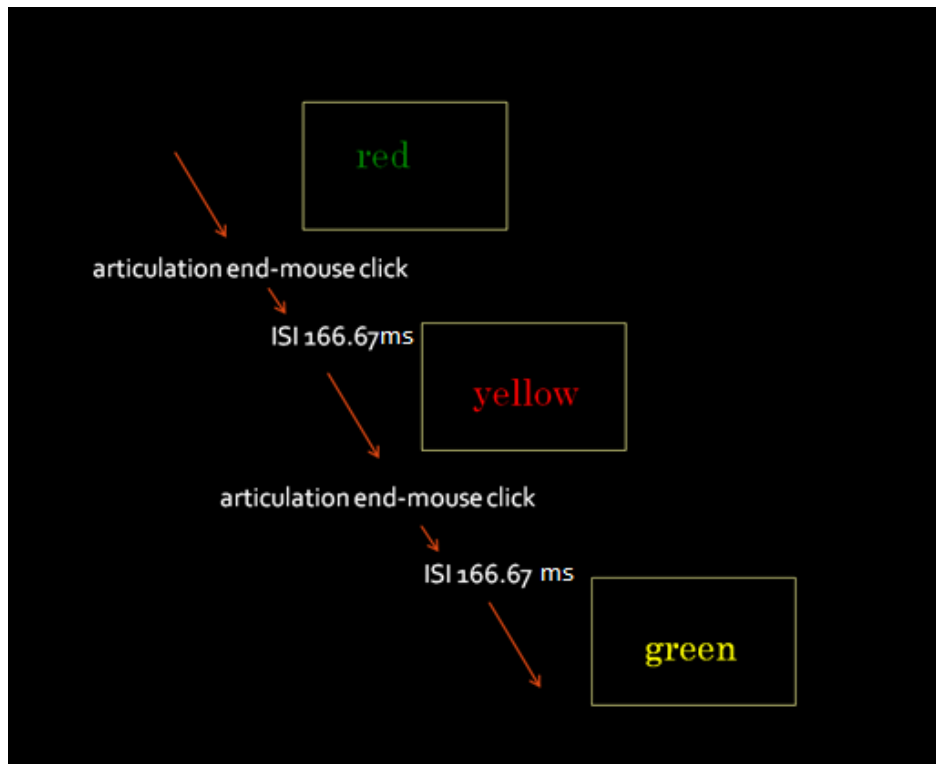


Figure 2.3. Illustration of the speeded task (SPEEDED-T).

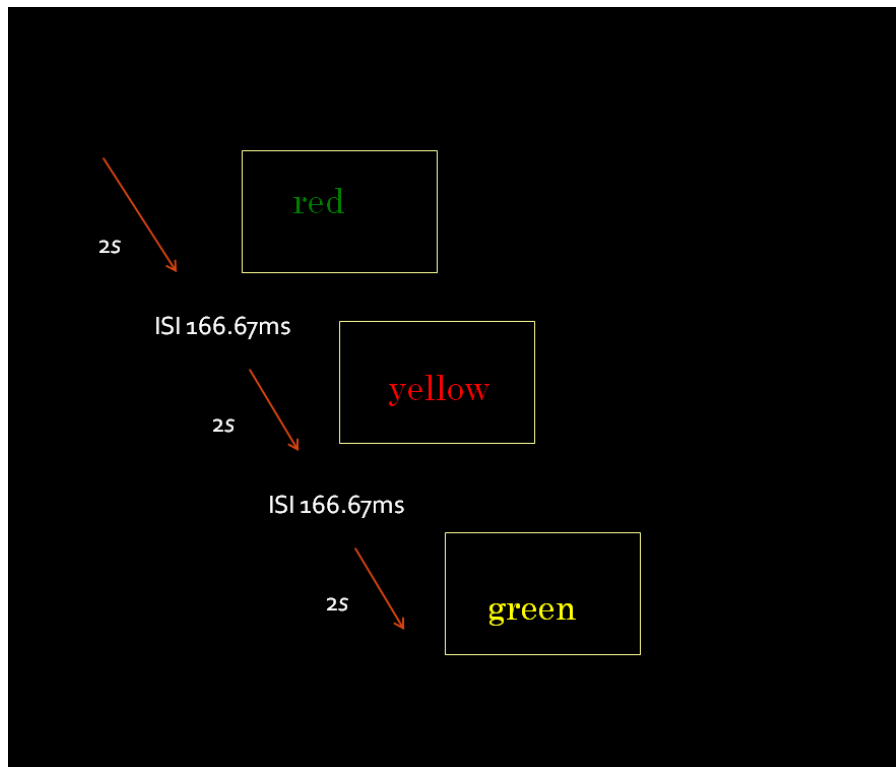


Figure 2.4. Illustration of the slow task (SLOW-T).

RAN. Two ran tests were used, including numbers (RAN-N) and colors (RAN-C). Each RAN task consisted of 50 items presented in five columns of ten rows. The participants were asked to name as quickly as possible the numbers or colors, avoiding errors, in an up-to-down fashion⁷.

Items of RAN-N were presented in white font on a black background. Each item extended 38×38 pixels on the screen. The RAN-N included 2, 3, 5, 7, and 9, pronounced /ðio/, /tria/, /pendε/, /εfta/, /εña/, respectively; these words were bisyllabic, including three with penultimate-syllable stress and two with final-syllable stress (*Appendix B*).

RAN-C items were colored rectangles on a black background, and extended 38×79 pixels on the screen. The colors used were the Greek words for red, green, yellow, blue, and brown (RGB#FF0000, #00FF00, and #FFFF0, #0070C0, #933300, respectively). The two additional colors (i.e., blue and brown) were chosen because

⁷Although RAN is usually introduced in a left-to-right fashion, we chose to adopt an up-to-down format to allow straightforward comparison with the card version of the Stroop task.

they are of similar written token frequency with red, green, and yellow, and between them (blue: 24.73; brown: 34.03 per million, from IPLR; Protopapas et al., 2012), and are easily distinguishable (*Appendix B*).

Reading measurements

Four word reading measurements were administered. Two were in a serial and two in a single-trial format.

For serial reading, two cards including words were used. Words were presented in white 20-pt Arial form. Each card consisted of 50 items in five columns of ten rows, including bisyllabic and trisyllabic words. In the one serial reading task, participants were instructed to read the words as quickly as possible, avoiding errors, from left-to-right in rows, while in the other task the instructions were the same with the exception that reading should follow a top-to-bottom scanning direction (*Appendix B*).

Discrete reading included words and pseudowords. For both tasks, words appeared on the screen in white 40-pt Arial form for 2sec. In the single word reading 110 words were included. For single pseudoword reading, 46 stimuli were used (*Appendix B*).

Inhibition test

The inhibition task consisted of four pairs of the Greek letters Α, Λ, Σ, Κ, Ζ, Μ, Ν, and Χ in white Arial font on a black background (letter pairs: Α-Λ, Σ-Κ, Ζ-Μ, Χ-Ν). In the initial phase participants were instructed to place their index fingers on the corresponding letter of a pair (e.g., Α-Λ) on the keyboard. After that, the task began, and each letter appeared randomly on the screen, and participants were required to press the Α button, when the letter Α appeared on screen, and the Λ button, when the letter Λ appeared. After 6 stimuli, the initial phase stopped, and participants were

instructed to inverse their responses, meaning to press the L button, when seeing A, and the A button, when seeing L. This procedure was repeated for all four pairs.

Procedure

Tasks were administered in two sessions with a one-week delay between them. In the first session, SLOW-T, single word and pseudoword reading, and inhibition measurements were administered. The second session included SPEEDED-T, card version of the Stroop task, RAN tasks, and serial word reading. For all serial tasks, exemplars of the original tasks were presented to the participants for verification of instructions' understanding and responses. Both sessions had similar duration (approx.15 min). Responses to oral tasks in both sessions were recorded on the computer via a headset.

Analyses

For single trial tasks responses were examined with CheckVocal (Protopapas, 2007) to determine accuracy and placement of the timing marks. Response times for SPEEDED-T and SLOW-T were subsequently logarithmically transformed to bring their distribution closer to normal.

The total duration for each serial task in each condition was subsequently measured on the waveform using Praat (Boersma & Weenink, 2012).

To compare the total duration of SPEEDED-T and SLOW-T tasks in both conditions, paired-samples t-tests were conducted.

To examine the effects of SPEEDED-T and SLOW-T on color naming times, we used function lmer of the lme4 package (Bates, Maechler, & Bolker, 2012) in R (R Core Team, 2012). We employed linear mixed-effects models with maximal random structures (Barr, Levy, Scheepers, & Tily, 2013), that is, including random intercepts for participants as well as random slopes for time, condition, and their interaction. For response time, in R notation the model formula was specified as:

$$\log RT \sim \text{condition} * \text{task} + (\text{condition} * \text{task} | \text{subject})$$

Here, “condition” refers to incongruent vs. control, and “task” to SPEEDED-T vs. SLOW-T. This model was compared to a simpler one without the interaction.

For error rates, we employed generalized linear mixed-effects models with binomial responses modeled via a logit link (Dixon, 2008). Again, a maximal random structure was used.

For correlation analyses between tasks only response time and total naming times in all tasks were included.

Inhibition was calculated computing response times for reversed trials.

Results

The t-test for the incongruent condition of both tasks for duration revealed a statistically significant difference ($t(40) = -28.22, p < .001$), and the same was observed for the control conditions ($t(40) = -52.70, p < .001$). Both results indicate that the SPEEDED-T was indeed shorter than the SLOW-T.

Table 2.1 shows the naming times per column for each condition.

The employment of the linear mixed effect model revealed that there was no significant interaction between task and condition ($\chi^2 = 3.44, df = 1, p = 0.06$). In contrast, a significant difference between the two tasks and between conditions was observed. Specifically, the incongruent condition was slower than the control condition ($\beta = 0.20, t = 11.5$), as expected. Regarding the task, the results showed that SPEEDED-T was faster than SLOW-T ($\beta = 0.04, t = 3.00$).

For error rates, as with response times, the interaction between task and condition was not significant ($\chi^2 = 0.126, df = 1, ns$). The incongruent condition had significantly more errors than the control condition ($\beta = 2.26, z = 2.86, p < 0.01$). For task, there was no statistically significant difference between them ($\beta = 0.55, z = 0.613, ns$; Figures 2.5 and 2.6).

Table 2.1.

Color naming times and number of errors in each condition and task

TASK	Naming Times				Number of errors			
	Incongruent		Control		Incongruent		Control	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
SLOW	690.12	99.54	535.01	61.83	0.00	0.01	0.006	0.01
SPEEDED	633.70	92.54	510.67	54.58	0.01	0.03	0.02	0.03

Note: “SLOW” refers to SLOW-T and “SPEEDED” to SPEEDED-T

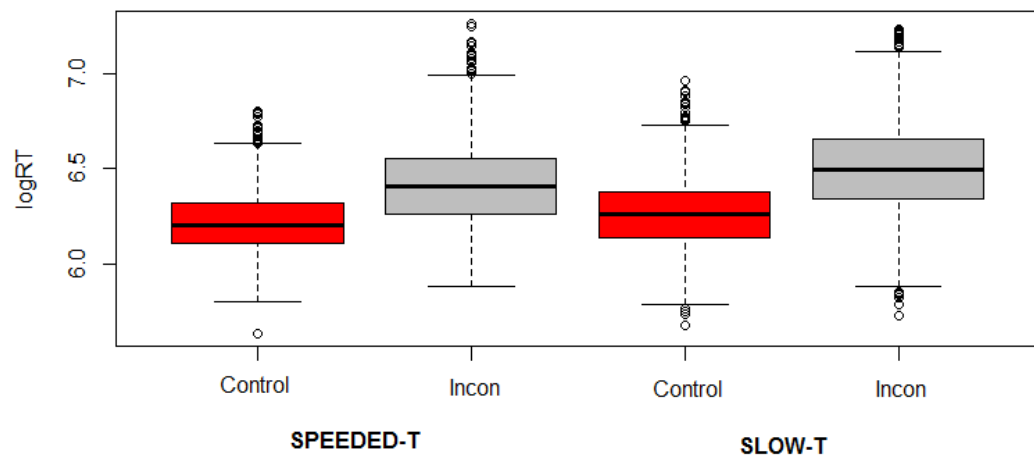


Figure 2.5. Color naming times in each task and condition

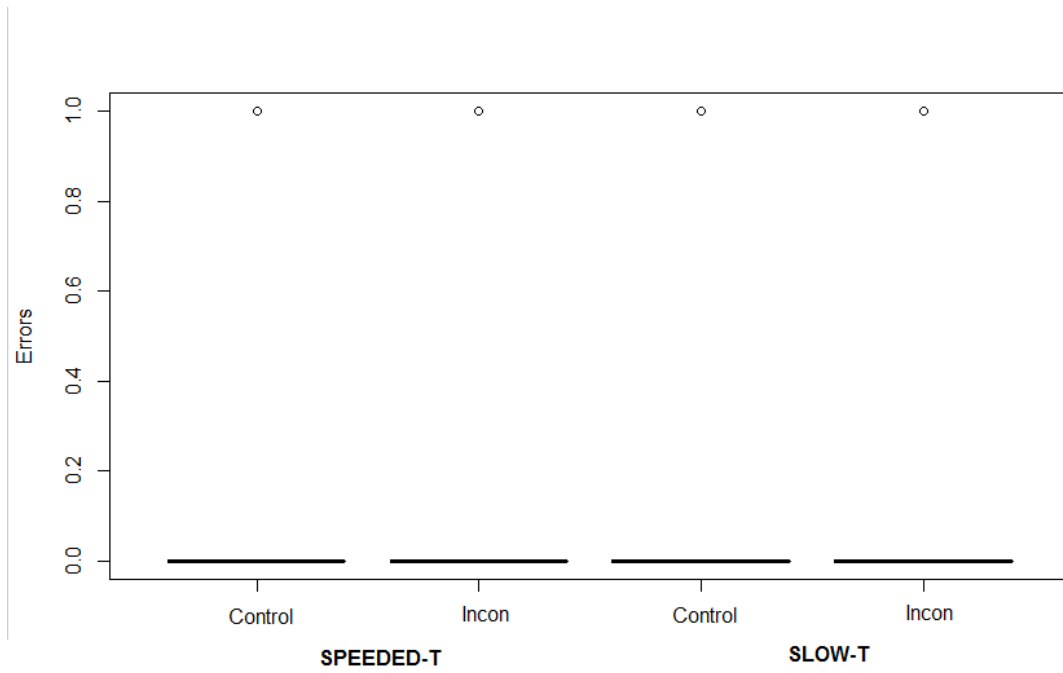


Figure 2.6. Errors in each task and condition.

Spearman's correlations followed (Table 2.2).

Table 2.2

Spearman's Correlations among tasks

	XXX-SL	INC-SP	XXX-SP	WORD	PSEUDO	LEFT-R	UP-D	INC-CAR	XXX-CAR	RAN-N	RAN-C	INHIB
INC-SL	.57***	.58***	.59***	.41**	.23	.14	.20	.51***	.22	.22	.14	.42**
XXX-SL		.43**	.62***	.64***	.34*	.30	.30	.47**	.37*	.30	.28	.59***
INC-SP			.54***	.42**	.10	.12	.18	.49**	.33*	.18	.28	.28
XXX-SP				.55***	.30	.28	.31*	.61***	.54***	.48**	.36*	.47**
WORD					.45**	.34*	.43**	.35*	.26	.35*	.18	.54***
PSEUDO						.35*	.39*	.13	.26	.27	.17	.63***
LEFT-R							.87***	.17	.43**	.58***	.25	.21
UP-D								.18	.36*	.58***	.27	.13
INC-CAR									.54***	.38*	.51***	.41**
XXX-CAR										.51***	.64***	.42**
RAN-N											.20	.28
RAN-C												.18

Note: "INC-SL"= incongruent condition of SLOW-T; "XXX-SL": control condition of SLOW-T; "INC-SP"= incongruent condition of SPEEDED-T; "XXX-SP": control condition of SPEEDED-T; "WORD": discrete word reading; "PSEUDO": discrete pseudoword reading; "LEFT-R": serial reading in left-right scanning; "UP-D": serial reading in up-down scanning; "INC-CAR": incongruent condition of card version of Stroop; "XXX-CAR": control condition of card version; "RAN-N"= RAN numbers; "RAN-C"= RAN colors, "INHIB"= inhibition task.

* $p < .05$. ** $p < .01$. *** $p < .001$.

For both conditions of the single trial tasks (i.e., SLOW-T, SPEEDED-T), discrete word/pseudoword reading, serial reading (left-right/up-down), RAN-C, RAN-N, and inhibition the results showed that the control condition of SPEEDED-T was positively correlated with discrete word reading, serial up-down reading, RAN-C, RAN-N, and inhibition. Correlation with pseudoword reading was only marginally significant. The rest of the correlations did not reach significance. The incongruent condition of SPEEDED-T was positively correlated with only one measure, namely discrete word reading. There was, also, a low positive correlation with RAN-C, although not significant. All other correlations were nonsignificant.

Regarding SLOW-T and the control condition, it was found to be positively correlated with discrete word reading, pseudoword reading, serial left-right and up-down reading, RAN-N, and inhibition. There was no significant correlation with RAN-C. The incongruent condition of SLOW-T was positively correlated with discrete word reading and inhibition. No other correlation reached significance. To sum up, the correlation results of SPEEDED-T and SLOW-T with the rest measurements indicate that the slowest version of the Stroop task was correlated with more reading and naming measurements than the SPEEDED-T, and, especially, in reference with the control condition of SLOW-T that was found to be positively correlated with all measurements. Surprisingly, the control conditions of both tasks were more highly correlated with inhibition than the incongruent condition of the tasks, which is generally thought to be more related to executive functions, and to inhibition specifically.

The results of correlation analysis between inhibition and the other measurements revealed that discrete word and pseudoword reading were positively related with inhibition. The correlations with the serial reading tasks, was nonsignificant. This was

also true for RAN-C and RAN-N. In contrast, both conditions of the card version of the Stroop task were positively related with inhibition.

As we stated in the introduction, one of our goals was to examine the relationship between the card version of the Stroop task with RAN-C and reading. Even though the control condition of the Stroop task was positively correlated with RAN-C, it was only moderately. The same pattern was observed for the incongruent condition, although somehow diminished. In relevance to the reading tasks, the results indicated that the serial word reading tasks were positively correlated with the control condition, but not with the incongruent condition. For the discrete reading tasks, the control condition correlated with neither task, whereas the incongruent condition correlated significantly and positively only with word reading. In contrast, RAN-C did not correlate significantly with reading measurements.

Since Study 1 revealed, that, during the card version of the Stroop task, performance does not remain stable, it was reasonable to assume that correlation of the Stroop task with reading measurements and inhibition may be altered, as the task progresses. So, we correlated those measurements with each column of the task for both conditions (Table 2.3).

Table 2.3.

Spearman's Correlations among columns of the card version of the Stroop task and inhibition, and reading measurements.

	INHIB	LEFT-	UP-D	WORD	PSEUDO
	R				
INC-1	.41**	.06	.06	.27	.15
INC-2	.30	.20	.27	.49**	.14
INC-3	.23	.15	.15	.18	.01
XXX- 1	.49**	.48**	.36*	.25	.30
XXX- 2	.43**	.38*	.30	.22	.26
XXX- 3	.39*	.46**	.40*	.39*	.26

Note: "INC-1": first column of the incongruent condition, "INC-2": second column of the incongruent condition, "INC-3": third column of the incongruent condition, "XXX-1": first column of the control condition, "XXX-2": second column of the control condition, "XXX-3": third column of the control condition.

* $p < .05$. ** $p < .01$.

For inhibition, the results showed that only the first column of the incongruent condition correlated significantly with inhibition. For the control condition, the first column correlated higher with inhibition than the second and third. For serial word reading, neither column of the incongruent condition correlated significantly with left-right reading, although an increasing tendency was observed. The same pattern was found for up-down reading. For discrete reading, pseudowords did not correlate significantly, but, interestingly, for word reading it was the second column that correlated significantly. For the control condition and serial reading, it was found that

the first and third column correlated more strongly with left-right reading than the second column, and the same was observed for up-down reading. For discrete reading, correlation of all columns with the pseudowords was nonsignificant. For discrete word reading and the control condition, there was no significant correlation for the first two columns, but the correlation was significant for the third column. To sum up, what was observed was that, as the task progressed, correlation with inhibition became weaker, but the opposite seems to be true for the reading measurements.

Because in this study, the RAN-C used did not follow the usual left-to-right format we were interested to examine, if this could be a possible reason for the nonsignificant correlation results regarding reading. So, we implemented Spearman's correlation analysis for RAN-N and reading measurements. RAN-N was positively correlated with both serial reading tasks, and with discrete word reading. Correlation with pseudowords was not significant. Moreover, there was a high positive correlation between the two serial word reading tasks indicating that there is no substantial difference between them, and, consequently, task format, regarding potential scanning, does not seem to be the crucial factor.

Discussion

The aim of the present study was to test two hypotheses regarding the pause time in RAN tasks. Wolf and Bowers (1999) proposed that pause time reflects inhibition processes, in the sense, that it is the time needed by participants to inhibit the previous item, and start processing the next. In contrast, Neuhaus et al. (2001) suggested that pause time is connected to the speed of processing, indicating retrieval. Our findings contradict the first, but support the second.

We have suggested that, if inhibition processes are reflected in pause time, like Wolf and Bowers (1999) propose, then response time in a speeded task, where extra

time is absent, should be greater compared to a slower task, where participants have more time to inhibit the previous item. This was not the case. In fact, response times in the speeded task were faster than in the slow task, indicating that the lack of extra time not only did not impair performance, but, in contrast, improved it. In other words, participants took advantage of the speed of the task. In addition, the slow task correlated more highly with reading and serial measurements than the speeded task.

Furthermore, the slow task was more highly related with inhibition, although the difference was not substantial. If pause time reflects inhibition processes, then the speeded task should have correlated more strongly with it than the slow. Again this was not the case. Participants inhibited to the same amount in both tasks and in both conditions. In addition, inhibition correlated significantly only with the discrete reading measurements, and not with serial (RAN-N and RAN-C included), suggesting that inhibition processes are not involved in serial reading, and are not necessary for naming in serial tasks.

Generally, our findings seem to support the retrieval hypothesis, however only partially. We have assumed that, if retrieval processes are to be reflected in pause time, then response times between tasks should not differ, but they did with the speeded task being faster. Moreover, tasks correlated differentially with reading measurements and serial tasks, suggesting that speed of processing may not determine pause times.

To my view, it is plausible to assume that retrieval is not the critical aspect in naming tasks, and even not necessary for execution of these tasks. RAN tasks have a defined target set of stimuli (usually five), which are presented prior to the task for verification of responses, and alternate during the task. If we take that as graded, the inevitable question is why participants need to retrieve each item and all lexical information in each response, since these items can be held in working memory. If so,

then pause time may reflect identification and verification of the correct response sustained in working memory. This is in line with Kirby's et al. (2010) suggestion that pause time involves attention and recognition of the items, although they acknowledge inhibition processes as well.

Our proposal is strengthened if we take into account the study of Amtmann, Abbott, and Berninger (2007), who examined the impact of phonological memory on RAN and RAS tasks in children and adults with dyslexia. They measured the time needed to complete each of ten rows in both tasks, and found that time increased as the rows progressed, and examined this course in reference to executive functions (inhibition) and phonological working memory. They concluded that inhibition and phonological working memory could predict the course of naming time.

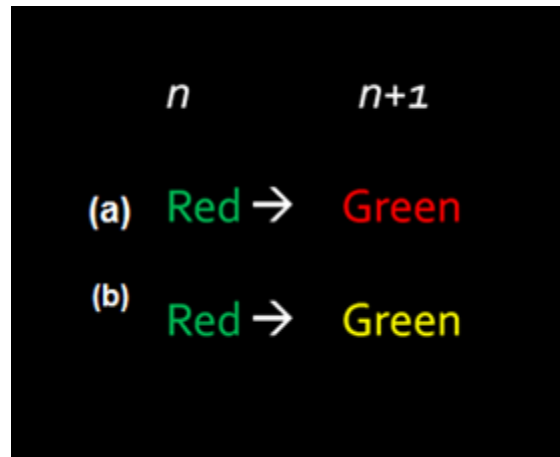
Although this study seems to support only our claim for the importance of working memory in RAN, and contradict the notion for no inhibition involvement, this is not the case. Interestingly, the authors used the card version of the Stroop task to measure inhibition. So, three points should be addressed. First, as we have seen in our study, card version of the Stroop task, RAN, and reading ability are positively correlated, so an improved performance in the Stroop task indicates improved performance to other tasks. This is also supported by other studies (Faccioli et al., 2008; Kapoula, et al., 2010; Protopapas, et al., 2007). Second, as we have reported in Study 1, decrements in performance are also observed in the card version of the Stroop task. Third, although the Stroop task was positively correlated to inhibition, reading measurements and RAN tasks were not. Taking these observations into account, it is plausible to assume that the Stroop task measures additional cognitive processes, despite inhibition, among them identification and selection of concepts in working memory.

One objection that could be raised against our overall interpretation, and the absence of inhibition processes in serial word reading and RAN tasks, is that in our experiment we implemented a single-trial inhibition task, in the sense, that one stimulus appeared on the screen each time, and, consequently, it did not capture inhibition processes present in RAN tasks and serial reading. In other words, it was the difference in task formats that produced the pattern observed. Although this point seems reasonable, it could not explain why the same task captured inhibition processes involved in the card version of the Stroop task. Importantly, this was not only true for the incongruent condition of the task, which is supposed to put more executive demands to participants, but also for the control condition. Interestingly, it was also the control condition of all Stroop tasks, and not RAN-C, that correlated positively with reading.

To my view, the above observations enhance our claim that retrieval is not necessary for naming tasks, but selection and identification between competitive, highly activated responses in working memory are. To make this point more clear, two additional issues should be addressed.

First, it is generally thought that inhibition is time dependent, in the sense, that a stimulus response is inhibited for a limited time, and after that it is fully active again. Evidence is coming from literature examining the suppression effect. For the Stroop task, it is found that, when for example the word RED is written in green in trial n , reaction times are slower, when in trial $n+1$ the required response is “red” than unrelated, e.g., yellow (Figure 9.1). It is assumed that, because the correct response on trial n requires the suppression of “red”, it could not be fully active again in trial $n+1$, which leads to time cost. Two additional findings regarding this effect are that strict accuracy instructions are necessary for the effect to occur. Under speed emphasis the effect is absent. Second, the effect diminishes, as the time between trials increases,

but, paradoxically, the overall response times to all trials also increases in the largest stimulus onset asynchronies (Neil & Westberry, 1987; Salo, Henik, Nordhal, & Robertson, 2002).



*Figure 2.7.*Suppression (negative priming) effect. Trials in (a) produce slower reaction times compared to (b). The effect seems to attenuate as time between stimuli increases, but response times for both (a) and (b) increase as a function of time.

On the other hand, evidence coming from studies using EEG, studying error related negativity related to ACC, propose that during a task, like the Stroop task, a procedure is active, namely response checking, where participants compare the intended response with the actually produced response. Under moderate and not severe time pressure the Ne component is larger, in correct trials as well, and this is also true if instructions emphasize accuracy (Falkenstein, Hoorman, & Hohnsbein, 2000). As the authors state: “One possible alternative hypothesis could be that the Ne does not reflect the outcome of the above mentioned comparison process (i.e., error detection), but the ongoing process (i.e., response checking) itself, since response checking is necessary also in correct trials” (p. 101).

In reference to our experiment, this would mean that in the slow version of the Stroop task, there is more time for response checking to occur compared to the speeded version, and to the speeded version compared to the card version. In the slow task representations in working memory are activated *fully*, due to response checking, and need more time to be inhibited compared to the speeded version and the card

version. So, inhibition is more related to the slow version than the speeded and the card, and response times are slower. This process may be also present in discrete word and pseudoword reading, where time is not limited, and actual and intended responses are kept into working memory for checking.

Keeping in mind that inhibition is not related to naming tasks and reading measurements, the above suggest that their connection is due to identification and selection of the response. If selection of a response is impaired, due to high activation of competitors, then identification and selection becomes harder and delays, leading, presumably, to enlarged pause times. This is partially in agreement with the suggestion of Amtmann et al. (2007) that “inhibition is impaired by accumulating name codes from prior trials, making it more difficult to focus on the relevant name at the moment for the most immediate orthographic symbol” (p.808), with the difference that, to my view, it is identification and selection that are impaired.

This seems to be supported by the finding that, although the correlation with inhibition during the card version of the Stroop task became weaker, the correlation with reading measurements, and, especially, for the control condition became stronger. Moreover, taking into account the main conclusion of Study 1 that participants get depleted, as the task progresses, and color naming time increment is observed, this could mean that identification and selection of the correct response becomes harder, leading, presumably, to increased pause times, and, ultimately, to enhanced correlation of the task with reading. This idea does not contradict our conclusions made in Study 1, but they expand it. It seems that multiple interferences in the card version of the Stroop task, impair naming of the target, because they enhance alternative responses, resulting to an impairment of identification and selection of a response between highly activated competitors. Conflict monitoring, as

proposed in Study 1, does not exclude response checking, but includes it. Response checking is by its nature a conflict manifestation (Carter et al., 2007).

The above proposals are also supported by studies using eye-tracking. Specifically, Jukka and Olson (1995) examined, how word-length and word-frequency determines fixation duration in skilled readers and readers with dyslexia, and found that longer gaze duration was observed in both groups for low-frequency and long words, indicating a difficulty in identifying a word, if it was rare. Word identification difficulties and context-use for word recognition in dyslexia have also been observed in other studies (Bruck, 1993).

Recall that RAN tasks correlating strongly with reading ability are those that use highly familiar and interconnected, related stimuli derived from a limited reservoir of stimuli (e.g., letter [A –Z], numbers [1-10], colors [usually, blue, green yellow, red, black],but not objects which can include a variety of nondepended stimuli). This raises the question of the impact of competition on them, and consequently, reading. To my view, our findings suggest that word neighborhood and frequency may have an impact on the relationship of reading and naming time, in the sense, that, although competitive responses are active, a correct one has to be chosen. Our experiments suggest that it depends on the target set, and, if so, on the number of competitors. Further research has to address, if the relationship between RAN and reading stands for all words irrespective of frequency and neighborhood, and, if so, irrespective of syllable frequency and neighborhood. Inhibition processes may not play a crucial role, selection between competitors may.

Finally, Amtmann et al. (2007) have also proposed that habituation, due to repetitive naming of specific stimuli, may impact performance on RAN tasks, leading to gradually decline in naming times. This is supported by participants' self-reports that the words have loosed their meanings, and this is more evident in the Stroop task

were stimulus set size is smaller (three vs. five in RAN). This phenomenon is broadly known as semantic satiation, and refers to the subjective experience that after excessive exposure to a word an attenuation of the meaningfulness of the word occurs (Balota & Black, 1997). Although self-reports are vulnerable to criticism, and should be seen with caution, they are indicative of participants' state during the task. If as they claim, they feel that words have no meaning but continue to execute the task, this means that access to all lexical information of the words is not required. This point needs further investigation by modified tasks with different target sets, and their relationship to reading and phonological working memory.

In conclusion, main goal of the present study was to examine two alternatives regarding pause time in RAN tasks. Our results contradict the inhibition hypothesis, and support, partially, the retrieval hypothesis. We have proposed that, because of the nature of these tasks, identification and selection between equally activated concepts sustained in working memory may be reflected in pause time. To my view, this point requires further investigation with a variety of tasks.

*Appendix A***Figure A1.** Incongruent Condition of the Card Version.

πράσινο	πράσινο	κίτρινο
πράσινο	κίτρινο	πράσινο
κίτρινο	πράσινο	κίτρινο
κόκκινο	κόκκινο	πράσινο
κόκκινο	κίτρινο	κίτρινο
κίτρινο	κίτρινο	πράσινο
κόκκινο	πράσινο	πράσινο
κόκκινο	κίτρινο	κόκκινο
κίτρινο	κόκκινο	κίτρινο
κόκκινο	πράσινο	πράσινο
πράσινο	κόκκινο	κόκκινο
πράσινο	πράσινο	κίτρινο
κόκκινο	κίτρινο	πράσινο
κόκκινο	πράσινο	κίτρινο
κίτρινο	κόκκινο	πράσινο
κόκκινο	κίτρινο	κίτρινο
κόκκινο	κόκκινο	κίτρινο
πράσινο	πράσινο	κόκκινο
κόκκινο	κόκκινο	κίτρινο
κίτρινο	κόκκινο	πράσινο

Figure A2. Control Condition of the Card Version with XXX.



Figure A3. Control Condition of the Card Version with Dissimilar Letters (X or B or L).

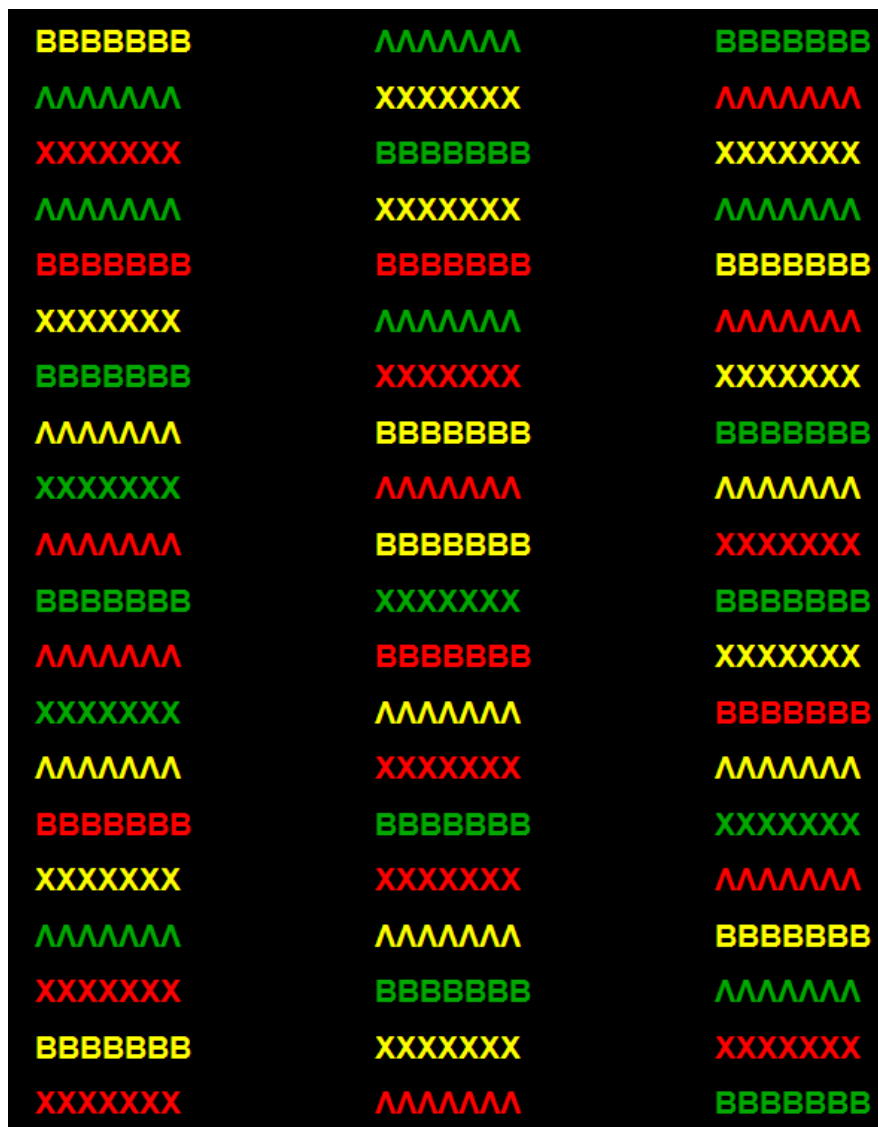
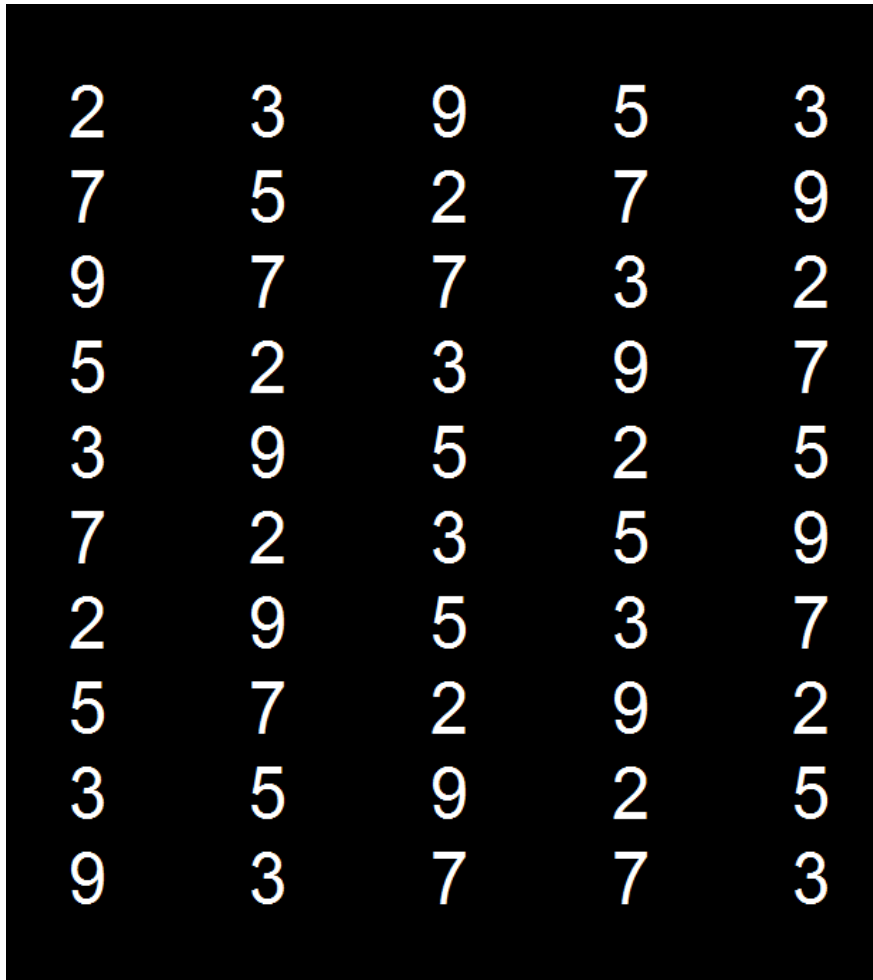


Figure A3. Control Condition of the Card Version with Animal Names.

ζαρκάδι	γορίλας	ζαρκάδι
γορίλας	φάλαινα	γορίλας
φάλαινα	ζαρκάδι	φάλαινα
γορίλας	φάλαινα	γορίλας
ζαρκάδι	ζαρκάδι	ζαρκάδι
φάλαινα	γορίλας	γορίλας
ζαρκάδι	φάλαινα	φάλαινα
γορίλας	ζαρκάδι	ζαρκάδι
φάλαινα	γορίλας	γορίλας
γορίλας	ζαρκάδι	φάλαινα
ζαρκάδι	φάλαινα	ζαρκάδι
γορίλας	ζαρκάδι	φάλαινα
φάλαινα	γορίλας	ζαρκάδι
γορίλας	φάλαινα	γορίλας
ζαρκάδι	ζαρκάδι	φάλαινα
φάλαινα	φάλαινα	γορίλας
γορίλας	γορίλας	ζαρκάδι
φάλαινα	ζαρκάδι	γορίλας
ζαρκάδι	φάλαινα	φάλαινα
φάλαινα	γορίλας	ζαρκάδι

Figure A5. Stimuli of the Single-item Version in all Conditions.

πράσινο	κόκκινο	πράσινο	κίτρινο
κίτρινο	κόκκινο	φάλαινα	γορίλας
ζαρκάδι	φάλαινα	γορίλας	ζαρκάδι
γορίλας	φάλαινα	ζαρκάδι	BBBBBBB
BBBBBBB	ΛΛΛΛΛΛ	ΛΛΛΛΛΛ	BBBBBBB
ΛΛΛΛΛΛ	XXXXXXX	XXXXXXX	XXXXXXX

*Appendix B***Figure B1.** RAN-Numbers

2	3	9	5	3
7	5	2	7	9
9	7	7	3	2
5	2	3	9	7
3	9	5	2	5
7	2	3	5	9
2	9	5	3	7
5	7	2	9	2
3	5	9	2	5
9	3	7	7	3

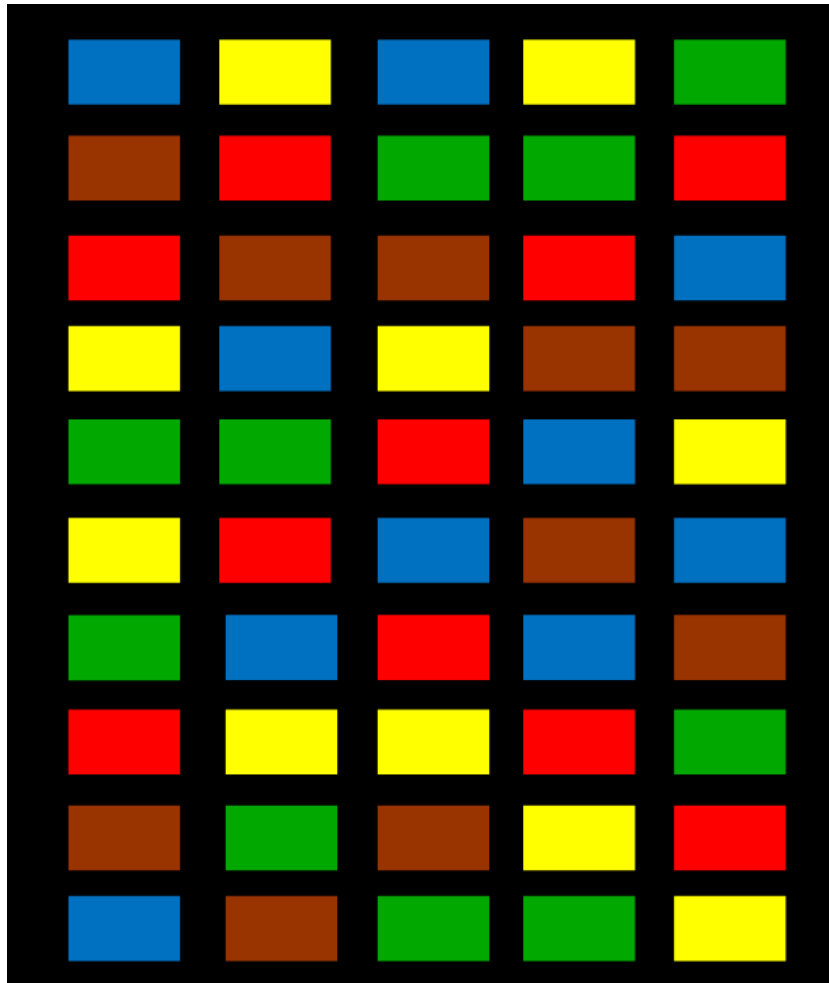
Figure B2. RAN-Colors

Figure B3. Serial Reading Measurements. (a) left-to-right; b) top-to-bottom.

(a)

ρόδου	μέση	πηγές	ρούχα	χωριό
μελών	πόλης	τέχνης	χάσμα	φλόγες
γενιά	εστία	χορός	βολή	φήμη
βλέμμα	πόδια	γνώσης	κουτί	ουρά
πείσμα	βουνά	ρεύμα	χαρτιά	φάσμα
μαθητών	βαλβίδα	εκπλήξεις	μηχανές	κερκίδα
ποτήρι	ορχήστρα	κατώφλι	δήμαρχοι	βιβλία
παπούτσια	δοκάρι	μαρμάρων	βενζίνη	μεμβράνη
λουλούδια	σύμπτωση	κουζίνα	φοιτητές	έκτασης
ποντίκι	εκπτώσεις	δίλλημα	μελέτες	φίλαθλοι

(b)

ταξί	λόγια	λάδι	συλλόγου	δάνειο
παιδιά	μέρη	αυγή	χειμώνα	σχήματα
στρώμα	τρύπες	φράγκα	λεξικό	στρεμμάτων
ξύλο	ρόλος	όψη	κατοίκους	οδηγού
στιγμή	νύχτα	δάφνη	παιχνιδιών	φαγητό
χώρου	ρυθμούς	συντροφιά	τιμόνι	θάλασσας
δόντια	σκάλα	προόδου	εισόδου	αρχηγό
φορές	μισθών	συντάκτης	συστάσεις	σύνθεση
πάρκο	ήχου	φορτηγά	πίνακα	πακέτου
δείγμα	οδός	πρόσκληση	σκουπίδια	καυσίμων

Figure B4. Discrete Reading. (a) Stimuli of word reading; (b) Stimuli of pseudoword reading.

(a)

αναγκάστηκε	επανάσταση	ευχαριστημένος
εικασία	διηγήθηκε	ανακούφιση
εμπιστοσύνη	λεπτομέρειες	εργοστάσιο
υπολοχαγός	συμπεριφορά	χαμογελώντας
πανεπιστήμιο	συμπλήρωσε	προσπαθούσε
παντρευτεί	μαθηματικά	περισσότερο
περιουσία	ιδιαίτερα	αντίρρηση
επρόκειτο	συγκίνηση	ατμόσφαιρα
ευκαιρία	τουλάχιστον	επίσκεψη
συνάντηση	χαιρέτησε	οργάνωση
διάρκεια	χρειάζεται	απόγευμα
απέναντι	εκκλησία	αργότερα
καφενείο	απάντησε	σημασία
στράφηκε	έκφραση	ευχαριστώ
κατάσταση	βουλευτής	έκπληξη
πρόεδρος	εξαπίας	δεσποινίς
αίσθηση	πιστεύω	αντάρτες
σκέφτεται	βρίσκονταν	καθρέφτη
κλαίγοντας	άγαλμα	ανθυπολοχαγός
συνταγματάρχης	λογαριασμό	τριαντάφυλλο
απογοήτευση	ενθουσιασμό	δευτερόλεπτα
απελπισία	εμφανίστηκε	γερουσιαστής
συλλογίστηκε	πραγματικότητα	χειρονομία
κατευθείαν	φωτογραφία	αισθάνθηκε
αποκρίθηκε	αγκάλιασε	απαντήσει
μεσάνυχτα	υπουργείο	κυβέρνηση
διευθυντής	συντροφιά	συγγνώμη
χρειάστηκε	συνέβαινε	συγγενείς
ταυτόχρονα	μουρμούρισε	αποτέλεσμα
στρατιώτες	διοικητής	περίπτωση
ολόκληρη	ενδιαφέρον	καινούργιο

τελευταίο	εντύπωση	σταμάτησε
αυτοκίνητο	συζήτηση	άνθρωπος
πράγματα	διάστημα	γειτονιά
ασφάλεια	πρόκειται	επιτέλους
συνέχεια	εικασία	έμοιαζε
ένιωθε	φαινόταν	χρήματα

(b)

ετραδόκνια	προχλίο	επαπλιψός
ταπλακή	επασλόβουν	ψέρτανο
φεκτονίξοσι	κλαξεδές	ππαρασβότος
οποστανάς	σχονειάπανε	μετόσφαλη
λοσοσθάση	σπόλωρας	πρακτεοφρίσο
οκέπλη	ανησάπεις	εκταστεήφας
ππορχανού	οντέπλοτα	αξόβλοξη
πλησμόσται	ρεζαπόφεςμα	χεφταψίνετο
απαστεβέκι	ηντοκλωχή	σμονεχρόα
πλαρυμπόναν	δενοτροτεκό	ορκελικτάστε
σenaξπροδόν	ναστεόλμιχνα	σοστούχα
βαρνεσκήχτορο	κεσμακρηντός	μοσοχλοτής
εξόρεσλη	φαχταψοταρί	ξιφταμοχερτό
πρακλεσταρτικές	λακεχαρμίο	νιοψαρτασμάρος
φπιψάχου	δεαρδαμπά	σπαλεστοκή
οκωπτισά		

References

- Amtmann, D., Abbott, R. D., & Berninger, V. W. (2007). Mixture growth models of RAN and RAS row by row: Insight into the reading system at work over time. *Reading and Writing, 20*, 785–813.
- Anderson, S. W., Podwall, F. N., & Jaffe, J. (1984). Timing analysis of coding and articulation processes in dyslexia. *Annals of the New York Academy of Sciences, 433*, 71–86.
- Balota, D. A., & Black, S. (1997). Semantic satiation in healthy young and older adults. *Memory & Cognition, 25*, 190-202.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.
- Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. <http://CRAN.R-project.org/package=lme4>.
- Baumeister, R. F. (2002). Ego depletion and self-control failure: An energy model of the self's executive functions. *Self & Identity, 1*, 129–136.
- Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science, 16*, 351–355.
- Blachman, B. A. (1984). Relationship of rapid naming ability and language analysis skills to kindergarten and first-grade reading achievement. *Journal of Educational Psychology, 76*, 610–622.
- Boersma, P., & Weenink, D. (2012). *Praat: Doing phonetics by computer* (Version 5.3.17) [Computer program]. Retrieved from <http://www.praat.org>.
- Boeksem, M. S. M., & Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain Research Reviews, 59*, 125-139.

- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624-652.
- Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*, 539-546.
- Botvinick, M. M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, *402*, 179-181.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*, 177-178.
- Bruyneel, S. D., Dewitte, S., Francis, P. H., & Dekimpe, M. G. (2009). I felt low and my purse feels light: Depleting mood regulation attempts affect risk decision making. *Journal of Behavioral Decision Making*, *22*, 153-170.
- Denckla, M. B., & Cutting, L. E. (1999). History and significance of rapid automatized naming. *Annals of Dyslexia*, *49*, 29-42.
- Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the on-line monitoring of performance. *Science*, *280*, 747-749.
- Carter, C. S., Mintun, M., & Cohen, J. D. (1995). Interference and facilitation effects during selective attention: An H₂¹⁵O PET study of Stroop task performance. *NeuroImage*, *2*, 264-272.
- Carter, C. S., & van Veen, V. (2007). Anterior cingulate cortex and conflict detection: An update of theory and data. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 367-379.
- Cohen, J. D., & Huston, T. A. (1994). Progress in the use of interactive models for understanding attention and performance. In C. Umiltà & M. Moscovitch

- (Eds.), *Attention and performance XV* (pp. 453-456). Cambridge, MA: MIT Press.
- Cobbold, S., Passenger, T., & Terrell, C. (2003). Serial naming speed and the component elements of speech time and pause time: Relationships with the development of word-level reading in children aged four to five years. *Journal of Research in Reading, 26*, 165–176.
- Dayan, P., & Solomon, J. A., (2010). Selective Bayes: Attentional load and crowding. *Vision Research, 50*, 2248- 2260.
- Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language, 59*, 447–456.
- Dulaney, C. L., & Rogers, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 470–484.
- Egner, T., & Hirsch, J. (2005). The neural correlates and functional integration of cognitive control in a Stroop task. *NeuroImage, 24*, 539-547.
- Eriksen, B.A., & Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143-149.
- Eriksen, C.W., Pan, K., & Botella, J. (1993). Attentional distribution in visual space. *Psychological Research, 56*, 5-13.
- Everatt, J., Warner, J., & Miles, T. R. (1997). The incidence of Stroop interference in dyslexia. *Dyslexia, 3*, 222–228.
- Faccioli, C., Peru, A., Rubini, E., & Tassinari, G. (2008). Poor readers but compelled to read: Stroop effects in developmental Dyslexia. *Child Neuropsychology, 14*, 277–283.
- Falkenstein, M., Hoormann, J., Crist, S., & Hohnsbein, J. (2000). ERP components in reaction errors and their functional significance: A tutorial. *Biological*

- Psychology*, 51, 87-1007.
- Felton, R. H., Naylor, C. E., & Wood, F. B. (1990). Neuropsychological profile of adult dyslexics. *Brain and Language*, 39, 485–497.
- Gatti, S. V., & Egeth, H. E. (1978). Failure of spatial selectivity in vision. *Bulletin of the Psychonomic Society*, 11, 181-184.
- Georgiou, G. K., Parrila, R., & Kirby, J. (2006). Rapid naming speed components and early reading acquisition. *Scientific Studies of Reading*, 10, 199- 220.
- Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 875-894.
- Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, 118, 13-42.
- Govorun, O., & Payne, B. K. (2006). Ego-depletion and prejudice: Separating automatic and controlled components. *Social Cognition*, 24, 111–136.
- Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin*, 136, 495–525.
- Henik, A., Bibi, U., Yanai, M., & Tzelgov, J. (1997). The Stroop effect is largest during first trials. *Abstracts of the Psychonomic Society*, 2, 57.
- Inzlicht, M., & Gutsell, J. (2007). Running on empty: Neural signals for self-control failure. *Psychological Science*, 18, 933–937.
- Jukka, H., & Olso, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1430- 1440.
- Kahneman, D., & Chajczyk, D. (1983). Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental*

Psychology: Human Perception and Performance, 9, 497-509.

- Kahneman, D., & Henik, A. (1981). *Perceptual organization and attention*. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 181-211). Hillsdale, NJ: Erlbaum.
- Kapoula, Z., Lê, T.-T., Bonnet, A., Bourtoire, P., Demule, E., Fauvel, C., Quilicci, C., & Yang, Q. (2010). Poor performance in 15-year old dyslexic teenagers. *Experimental Brain Research*, 203, 419–425.
- Kirby, J., Georgiou, G. K., Martinussen, R., & Parrila, R. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly*, 45, 341-362.
- Korohnen, T. T. (1995). The persistence of rapid naming problems in children with reading disabilities: A nine-year follow-up. *Journal of Learning Disabilities*, 28, 232–239.
- Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2011). Pupillary Stroop effects. *Cognitive Processing*, 12, 13-21.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109, 163–203.
- MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition*, 26, 201–211.
- MacLeod, C.M. (2005). The Stroop task in cognitive research. In A. Wenzel & D. S. Rubin (Eds.) *Cognitive methods and their application to clinical research* (pp. 17–40). Washington, DC: American Psychological Association.
- Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, 126, 247–259.

- Muraven, M., Rosman, H., & Gagné, M. (2007). Lack of autonomy and self-control: Performance contingent rewards lead to greater depletion. *Motivation and Emotion, 31*, 322–330.
- Neil, W.T. & Westberry, R.L. (1987). Selective attention and the suppression of cognitive noise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 327-334.
- Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and Reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology, 63*, 427- 452.
- Obregon, M. (1994). *Exploring naming timing patterns by dyslexic and normal readers on the serial RAN task*. Unpublished master's thesis, Tufts University, Medford, MA.
- Pardo, J. V., Pardo, P., Janer, K. W., & Raichle, M. E. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy of Science, USA, 87*, 256-259.
- Perret, P., & Ducrot, D. (2010). Viewing-position effects in the Stroop task: Initial fixation position modulates Stroop effects in fully colored words. *Psychonomic Bulletin & Review, 17*, 550-555.
- Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team (2014). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-117, <http://CRAN.R-project.org/package=nlme>.
- Pocheptsova, A., Amir, O., Dhar, R., & Baumeister, R. F. (2009). Deciding without resources: Resource depletion and choice in context. *Journal of Marketing Research, 46*, 344–355.
- Protopapas, A., Tzakosta, M., Chalamandaris, A., & Tsiakoulis, P. (2012). IPLR: An

- online resource for Greek word-level and sublexical information. *Language Resources & Evaluation*, 46, 449–459.
- Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, 39, 859–862.
- Protopapas, A., Archonti, A., & Skaloumbakas, C. (2007). Reading ability is negatively related to Stroop interference. *Cognitive Psychology*, 54, 251–282.
- R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Treisman, A. M., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Neuhaus, G., Carlson, C. D., Jeng, W., Post, Y., & Swank, P. (2001). The reliability and validity of rapid automatized naming scoring software ratings for the determination of pause time and articulation component durations. *Educational and Psychological Measurement*, 61, 490–504.
- Neuhaus, G., Foorman, B. R., Francis, D. J., & Carlson, C. D. (2001). Measures of information processing in rapid automatized naming (RAN) and their relation to reading. *Journal of Experimental Child Psychology*, 78, 359–373.
- Salo, R., Henik, A., & Robertson, L. C. (2001). Interpreting Stroop interference: An analysis of differences between task versions. *Neuropsychology*, 15, 462–471.
- Scarborough, H. S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia*, 48, 115–136.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.

- Tecce, J. J., & Dimartino, M. (1965). Effects of heightened drive (shock) on performance in a tachistoscopic color-word interference task. *Psychological Reports, 16*, 93-94.
- Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Science, 15*, 160- 168.
- Wolf, M., Bally, H., & Morris, R. (1986). Automaticity, retrieval processes, and reading: A longitudinal study of average and impaired readers. *Child Development, 57*, 988–1000.
- Wolf, M. & Bowers, P.G.(1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology, 91*, 415-438.