Acta Psychologica 148 (2014) 204–208

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Word reading practice reduces Stroop interference in children ☆

Athanassios Protopapas [a,b,*], Eleni L. Vlahou [c], Despoina Moirou [b], Laoura Ziaka [b]

[a] Department of Philosophy and History of Science, University of Athens, Greece
[b] Graduate Program in Basic and Applied Cognitive Science, University of Athens, Greece
[c] Institute of Psychology & Education, University of Ulm, Germany

ARTICLE INFO

Article history:
Received 2 October 2013
Received in revised form 23 December 2013
Accepted 14 February 2014
Available online xxxx

PsycINFO classification:
2340

Keywords:
Stroop
Practice
Reading
Naming
Automaticity
Training

ABSTRACT

Stroop interference is thought to index reading automaticity and is expected to increase with reading practice and to decrease with improved color naming. We investigated the effects of practice in word reading and color naming on interference in 92 adults and 109 children in Grades 4–5. For children, interference was reduced after reading practice with color words. In neither group was interference affected by practice in color naming of neutral stimuli. These findings are consistent with a direct negative relationship between reading ability and interference and challenge the automaticity account in favor of a blocking mechanism whereby interference is determined by the delay to inhibit the reading response rather than by the efficiency of color naming.

1. Introduction

Stroop interference is one of the most studied effects in cognitive psychology (MacLeod, 1991, 2005). The task appears deceptively simple: Participants must name the color in which color words are printed. When this is incongruent with the word, such as the word "red" printed in green, the correct response ("green") takes longer to produce, and is less accurate, compared to a neutral condition, such as naming the color of meaningless strings (e.g., "XXXX"). Interference is taken to index a conflict among two processing dimensions, namely word reading and color naming, both of which have the potential to produce a relevant response. One of these dimensions, namely word reading, is typically well-practiced and largely obligatory, whereas the other is not.

MacLeod (1991) reviewed theoretical proposals for this effect and concluded in favor of an automaticity hypothesis. Supported by connectionist modeling (Cohen, Dunbar, & McClelland, 1990), the automaticity approach posits that extensive practice in reading has caused the processing connections between the word input and the reading response to become stronger than the connections between the color input and the naming response. This difference causes interference, as processing through the weaker naming connections takes longer to overcome the dominant response arising from the stronger reading connections. A direct prediction of this theory is that practicing color naming should strengthen the corresponding connections and thereby reduce interference. In contrast, practice in word reading might increase interference, to the extent that further strengthening in this well-practiced skill is possible.

Alternative accounts emphasize the idea that verbal inputs are privileged relative to verbal outputs (e.g., Durgin, 2000, 2003; Glaser & Glaser, 1989, Sugg & McDonald, 1994; Virzi & Egeth, 1985). Along these lines, a recent computational approach based on a word production model posits a fundamental asymmetry between reading and naming, insofar as color naming must pass through concept activation whereas word stimuli achieve direct access to lemma retrieval and form encoding (Roelofs, 2003). In this model, color recognition activates a color concept, which leads to selection of the corresponding lemma. This two-step process disadvantages color naming, as it must always overcome a dominant word reading response that has been directly activated in the meantime. Production of the correct (color naming) response is possible only after the inappropriate (word reading) response has been blocked by an attentional mechanism. Therefore, interference is linked to the efficiency of the blocking mechanism and, potentially, to the speed of word reading: The incorrect response can only be blocked after it becomes available. A prediction of this theory is that practicing color naming should have a small or no effect on interference, because

☆ A partial preliminary report of this work appears in the Proceedings of the 35th annual conference of the Cognitive Science Society (Berlin, Germany, August 2013).

* Corresponding author at: MITHE, Ano Ilissia University Campus, GR-157 71 Zografos, Greece. Tel.: +30 210 727 5540; fax: +30 210 727 5530.
E-mail address: aprotopapas@phs.uoa.gr (A. Protopapas).

http://dx.doi.org/10.1016/j.actpsy.2014.02.006
0001-6918/

it is the waiting for blocking to take effect that primarily determines interference. In contrast, practicing word reading should *decrease* interference, via more rapid availability of the inappropriate responses, if reading performance has not already reached its maximum.

Despite many decades of efforts and a multitude of research reports, no study of the relationship of Stroop interference to practice in the individual dimensions that constitute it has been undertaken. The effects of isolated practice in neutral color naming or word reading on interference remain unknown. Previous studies have used mixed or incongruent-only stimuli and have succeeded to reduce (but not entirely remove) interference (e.g., Davidson, Zacks, & Williams, 2003; Dulaney & Rogers, 1994; MacLeod, 1998). The initial buildup of interference has also been achieved in a novel domain (shape naming) via training (MacLeod & Dunbar, 1988), offering support for a continuum of automaticity modulating interference. However, there was no study of the effects of training after interference was established.

The issue has attained renewed importance as recent studies have produced correlational evidence in favor of the blocking hypothesis and against the automaticity hypothesis. Specifically, Protopapas, Archonti, and Skaloumbakas (2007) found that children diagnosed with dyslexia, hence impaired in their reading, exhibited more interference than typically developing readers of the same age (see also Everatt, Warner, & Miles, 1997; Faccioli, Peru, Rubini, & Tassinari, 2008; Kapoula et al., 2010, for similar findings). Moreover, they found a negative association between reading and interference in the general population of children 11–14 years old, whereby better readers exhibited less interference.

Reverse interference, that is, a slowing down of word reading when the word is printed in an incongruent color, is not observed (Roelofs, 2003) except by atypical manipulations (MacLeod, 1991). However, under an automaticity account, it should be possible to create some reverse interference, at least temporarily, by focusing on the less practiced dimension. In elementary school children, before word reading has reached full automatization, intensive color naming practice might succeed in overcoming the efficiency difference and exert at least some effect. However, no study has examined the effects of practice on interference in school-age children.

In the present study the predictions regarding effects of dimensional practice on interference were tested directly. One group practiced color naming on the incongruent dimension, to quantify the extent of interference reduction by this manipulation. Two groups practiced the simple dimensions of word reading and color naming, using neutral stimuli, that is, words printed in white letters and colors of strings of Xs, respectively. Every group was tested for interference before and after practice, with as few trials as possible (determined in a pilot study), to minimize learning due to testing. Naming conditions were tested separately to minimize errors and focus on response times. A control group served as a reference baseline, tested twice without any practice. The experiment was carried out in two age groups: Educated adults are expected to have reached their potential in reading performance, therefore there should be no effects of reading practice. In contrast, from the automaticity point of view there is no reason to expect that color naming performance has been maximized, therefore under this account color practice should be effective. In contrast, children still have room for improvement in both word reading (evidenced in fluency norms) and color naming, therefore effects of practice should be potentially observable in every condition.

2. Method

2.1. Participants

The total sample included 92 adults 18–40 years old, mainly undergraduate and graduate students, and 109 children 9–11 years old attending Grades 4–5 from the general population. Both groups were as large as feasible given available resources. All participants were native speakers of Greek.

2.2. Materials

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 (33, 34, and 9 per million, respectively, from the IPLR; Protopapas, Tzakosta, Chalamandaris, & Tsiakoulis, 2012), and begin with voiceless stops, which facilitate response time triggering. The corresponding colors are familiar and easily distinguishable. The neutral word condition included these three words in white font.

Stimuli for the neutral color condition were made up of 7 repetitions of the letter X (no spaces) in red, green, and yellow colors (RGB #FF0000, #00FF00, and #FFFF00, respectively). For the incongruent condition the Greek words for red, green and yellow appeared in a non-matching color (i.e., one of the other two colors). All stimuli were presented on a laptop screen in 40-pt Arial font on a black background.

2.3. Procedure

For children, testing and practice took place under direct supervision by the experimenter in their school environment in a quiet room during class hours with the consent of parents, teachers, and school authorities. For adults, testing took place in a quiet room at the University or, occasionally, at home. Adult practice took place at home, unsupervised; to verify compliance, computer output from the practice runs was delivered to the experimenters in the last testing session.

2.3.1. Testing

On Day 1 and Day 5 interference measurements were taken (before and after practice, respectively). For the two color naming conditions, participants were asked to name the color of the ink as quickly as possible and to try to avoid errors. The neutral condition was administered first (24 stimuli, including 8 in each color), followed by the incongruent condition (24 stimuli, including 4 in each mismatching word-color combination). For word reading, participants were asked to read the word; neutral and incongruent conditions were made analogously to color naming. The word reading tasks were always presented after the color naming tasks, to avoid short-term effects on color naming interference. Each stimulus appeared on the screen for up to 2 s. Responses were recorded via a headset under the control of DMDX (Forster & Forster, 2003). Practice trials preceded data collection. The entire testing session lasted about 3 min.

2.3.2. Practice

Participants were assigned randomly into one of four conditions to practice for three consecutive days (Days 2–4). Group A practiced color naming of incongruent stimuli, Group B color naming of neutral stimuli (colored Xs), and Group C practiced word reading of neutral stimuli (words in white font). Practice trials were identical to the corresponding testing conditions, with a single stimulus appearing on the screen for up to 2 s. Children were required to complete one block of 144 trials per day. Adults completed one block of 192 trials per day. Group D simply waited out the 3-day period.

2.4. Analyses

Data were analyzed with mixed-effects models (Baayen, 2008; Baayen, Davidson, & Bates, 2008) with maximal random structures (Barr, Levy, Scheepers, & Tily, 2013) using function lmer of the lme4 package (Bates, Maechler, & Bolker, 2012) in R (R Core Team, 2012). For response time (RT) analysis, p values were derived by comparison (via likelihood ratio) to a model excluding the effect under consideration. Accuracy was analyzed with generalized mixed-effects models for binomial distributions (Dixon, 2008) via a logit transformation (Jaeger, 2008), using the same R function.

3. Results

Responses were examined with CheckVocal (Protopapas, 2007) to determine accuracy and placement of timing marks. RTs exceeding 200 ms, from correct responses only, were logarithmically transformed to bring their distribution closer to normal. Group-blind examination of participant accuracy and speed distributions led to exclusion of 12 children from color naming and 4 from word reading analyses, due to more than 11% errors. In the following analyses there were 21, 24, 27, and 25 children in groups A, B, C, and D, respectively, in color naming and 18, 28, 31, and 28 in word reading. No adults were removed, leaving 25, 23, 22, and 22 participants in both analyses.[1] Tables 1 and 2 list the mean response times and error rates per group for each task.

3.1. Effects on interference

The predicted hypotheses were examined in planned tests of interactions by group, against the reference (no-practice) group D. Specifically, interference change was tested as a triple interaction of group (A/B/C vs. D) × time (pre vs. post) × condition (incongruent vs. neutral). The effects of practice on color naming interference, relative to the control condition, are shown graphically in Fig. 1 for both participant groups.

3.1.1. Children

For Group A, who practiced color naming of incongruent stimuli, there was a significant reduction in color naming interference, both in RT ($\hat{\beta} = .14$, $t = 3.73$, $p < .001$) and in accuracy ($\hat{\beta} = 2.01$, $z = 2.30$, $p = .022$). There was no change in word reading interference (RT, $\hat{\beta} = .02$, $t = .65$, $p = .517$; accuracy, $\hat{\beta} = 2.58$, $z = 1.37$, $p = .169$).

In Group B, who practiced color naming of neutral stimuli, there was no change in color naming interference (RT, $\hat{\beta} = .01$, $t = .18$, $p = .859$; accuracy, $\hat{\beta} = .06$, $z = .07$, $p = .942$) or word reading interference (RT, $\hat{\beta} = .02$, $t = .74$, $p = .461$; accuracy, $\hat{\beta} = .77$, $z = .46$, $p = .648$).

In Group C, who practiced word reading of neutral stimuli, there was a significant reduction in color naming interference, for both RT ($\hat{\beta} = .08$, $t = 2.23$, $p = .029$) and accuracy ($\hat{\beta} = 1.46$, $z = 2.04$, $p = .041$). There was no change in word reading interference (RT, $\hat{\beta} = .05$, $t = 1.37$, $p = .173$; accuracy, $\hat{\beta} = .39$, $z = .27$, $p = .790$).

3.1.2. Adults

For Group A there was a significant reduction in color naming interference, only in RT ($\hat{\beta} = .13$, $t = 3.44$, $p = .001$; accuracy: $\hat{\beta} = .39$, $z = .33$, $p = .744$). There was no other effect in interference, for either color naming or word reading, in any group (all $p > .3$).

3.2. Reliability considerations

To partially address concerns due to interim analysis and the lack of power analysis, we examined random subsets of participants in a custom bootstrap procedure with 1000 group pairs sampled equally from each of the critical experimental groups (B, C) and the control group (D). For children, group size 10 resulted in $t > 2$ for 5% of the samples for Group B and 19% in Group C; the corresponding proportions for group size 15 were 3% and 21%. For adults, proportions were 5% and 2%, for group size 10, and 3% and 1%, respectively. Thus, it appears in this post-hoc analysis that the critical finding for word reading practice (Group C) survives four times more frequently than chance in underpowered groups whereas none of the other comparisons exceeded the expected proportion of 5%, lending some additional support to the main finding.

[1] Results are robust to inclusion of all participants as well as to more stringent (group-blind distribution-based) additional exclusion of 2 children, one in color and one in word analysis, due to mean RTs exceeding 1000 and 900 ms, respectively, as well as 4 adults who exceeded a 5% error or a 700-ms mean RT cutoff.

Table 1
Color naming response times (ms; top) and accuracy (percent error; bottom) per group.

Group	Before practice				After practice			
	Neutral		Incongruent		Neutral		Incongruent	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Adults								
A	527.9	62.0	678.9	107.8	517.6	75.3	573.9	100.2
B	545.5	53.4	668.7	86.7	539.0	74.0	626.2	73.7
C	528.9	66.8	665.4	84.6	524.1	80.8	641.4	123.5
D	588.0	88.3	721.5	96.2	580.7	85.8	702.4	97.5
Children								
A	694.5	78.1	910.2	136.0	785.1	125.8	881.6	134.4
B	633.2	80.0	814.9	83.3	680.8	105.1	875.4	95.3
C	690.3	122.4	917.6	164.2	731.3	111.4	893.2	136.8
D	744.0	120.7	952.0	135.4	757.6	117.3	973.0	141.7
Adults								
A	0.7	1.6	1.7	3.2	0.8	2.4	1.7	3.2
B	0.5	1.4	2.0	3.0	0.9	2.2	2.0	2.8
C	0.6	1.9	1.0	2.5	1.7	2.8	1.7	3.8
D	0.9	1.8	2.3	4.7	0.8	1.7	2.0	2.9
Children								
A	1.2	3.1	8.5	8.0	1.6	2.9	3.2	4.8
B	2.6	3.7	5.7	4.8	1.0	1.8	4.4	4.8
C	1.6	2.5	7.0	5.8	2.6	3.5	5.3	5.3
D	2.5	3.0	8.7	5.5	1.2	1.9	7.2	7.5

3.3. Effects on neutral stimuli

In an attempt to disentangle the statistical interactions indicating effects of practice on interference, we proceeded with post-hoc comparisons examining responses to the neutral stimuli in the two task dimensions before and after practice. Practice effects were analyzed as interactions by group, against the reference (no-practice) group D, to control for learning due to (pre-)testing.[2] For example, color naming change was tested as an interaction of group (A/B/C vs. D) × time (pre vs. post) on color naming responses in the neutral condition. Comparisons across dimensions were tested as a triple interaction of group (A/B/C vs. D) × time (pre vs. post) × task (color naming vs. word reading).

3.3.1. Children

For Group A, there was a significant increase in RT to both dimensions (color, $\hat{\beta} = .09$, $t = 3.34$, $p = .002$; word, $\hat{\beta} = .18$, $t = 4.88$, $p < .001$); this increase was greater for word reading ($\hat{\beta} = .01$, $t = 3.66$, $p = .001$). For Group B, there was a significant increase in word reading RT ($\hat{\beta} = .06$, $t = 2.11$, $p = .039$), not significantly different from the increase in color naming RT ($\hat{\beta} = .01$, $t = 0.41$, $p = .685$), which, however, was not significant ($\hat{\beta} = .05$, $t = 1.67$, $p = .100$). For Group C, there was a significant decrease in color naming accuracy ($\hat{\beta} = 1.33$, $z = 2.11$, $p = .035$); the increase in color naming RT was not significant ($\hat{\beta} = .04$, $t = 1.49$, $p = .140$) and was only marginally greater than the increase in word reading RT ($\hat{\beta} = .05$, $t = 1.70$, $p = .094$), which was barely negative and not significant ($\hat{\beta} = -.003$, $t = -0.09$, $p = .931$).

3.3.2. Adults

There were no significant practice effects for either color naming or word reading, or for the interaction between the two, in any group (all $p > .17$).

[2] However, these comparisons fail to control for participation in an uninteresting procedure, as participants in Group D did not undergo any repetitive practice. This may have differentially affected the enthusiasm with which participants in the four groups, especially children, carried out the post-practice tasks.

Table 2
Word reading response times (ms; top) and accuracy (percent error; bottom) per group.

Group	Before practice				After practice			
	Neutral		Incongruent		Neutral		Incongruent	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Adults								
A	481.2	77.2	490.3	66.3	473.5	74.6	496.8	82.8
B	494.1	71.9	503.5	87.1	480.1	58.1	501.2	64.9
C	470.8	44.6	484.3	67.0	452.7	52.4	464.7	54.0
D	519.5	90.6	524.0	88.9	515.4	75.4	522.7	87.1
Children								
A	619.1	69.7	660.4	94.6	795.4	125.4	829.7	175.1
B	591.9	76.2	617.1	91.4	666.9	105.1	674.2	104.8
C	629.4	101.7	661.6	132.8	666.1	129.2	692.4	127.4
D	692.7	122.4	754.0	138.9	729.5	113.3	750.6	125.3
Adults								
A	0.0	0.0	0.2	0.8	0.2	0.8	0.5	1.9
B	0.2	0.9	0.2	0.9	0.0	0.0	0.4	1.3
C	0.0	0.0	0.4	1.8	0.4	1.2	0.0	0.0
D	0.0	0.0	0.2	0.9	0.2	0.9	0.0	0.0
Children								
A	0.2	1.0	0.5	1.4	0.5	1.3	9.4	11.5
B	0.3	1.1	0.4	1.7	0.4	1.3	1.3	3.2
C	0.3	1.1	1.1	2.6	0.4	1.3	1.9	3.9
D	0.9	2.1	1.2	2.3	0.7	2.0	0.9	2.1

4. Discussion

The data showed that, for children, a few hundred trials of reading aloud the words meaning red, green, and yellow, presented in a white font, resulted in diminished interference in naming the colors of these words in color-incongruent display. In contrast, practice in naming the same colors did not affect interference. These findings are consistent with the blocking hypothesis and inconsistent with the automaticity hypothesis for Stroop interference. It seems that word reading expertise, rather than color naming expertise, constitutes the bottleneck causing interference, even though word reading is an easier, faster, and largely obligatory task.

Color naming is generally thought to be lacking automaticity. This is the oft-cited reason for Stroop interference, as it appears self-evident that there is unrealized potential in color naming to be developed by practice. Because color naming is the main dimension of interference, leading to the appropriate response, one might expect large effects of color naming practice. Yet none were observed in either children or adults.

Neutral word reading practice by children led to a reduction of interference. The effect size was comparable to that achieved by incongruent color naming practice (the difference in interference change between Groups A and C was not significant, RT: $\hat{\beta} = .06$, $t = 1.59$, $p = .119$; accuracy: $\hat{\beta} = 1.07$, $z = 1.19$, $p = .234$). This finding is consistent with a negative relationship between reading and interference (Protopapas et al., 2007). It calls for a reinterpretation of the relationship between age and interference (Comalli et al., 1962). Perhaps the increase in interference in early elementary grades indexes the emerging conflict between the conceptually mediated task of color naming and the newly acquired skill of word reading. Subsequently, increasing expertise in reading is associated with a reduction in interference. Improvements in reading performance may partially account for these longitudinal effects.

Indirect evidence for practice effects on response potency was obtained in children's Groups B and C, consistent with the practiced dimension. Specifically, training with neutral stimuli on one dimension seems to have protected this dimension from increasing latencies or errors at post-test. These effects cannot explain the effects of practice on interference without recourse to the blocking hypothesis, because the relatively strengthened color naming responses in Group B were not accompanied by a reduction of interference.

Practice in color naming of incongruent words led to a reduction of interference in both groups. For adults, this finding is consistent with previous reports (Davidson et al., 2003; Dulaney & Rogers, 1994;

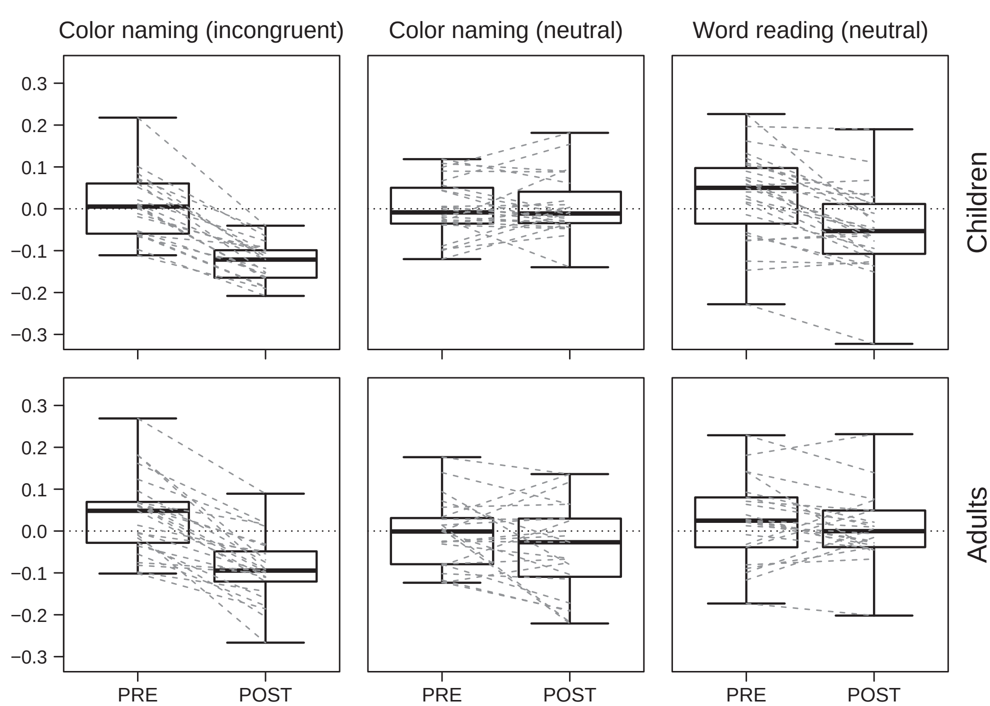

Fig. 1. Differences in color naming interference (i.e., the difference in log response times between the incongruent and neutral conditions) between each practice group (A: color naming of incongruent stimuli, B: color naming of neutral stimuli, C: word reading of neutral stimuli) and the corresponding control group (D: no practice) before (PRE) and after (POST) practice in each condition. The horizontal dotted line at zero indicates the baseline (i.e., the interference of the control group). Boxes enclose the middle 50% of the data; error bars extend to the full range. Dashed lines from pre to post join the performance of individual participants before and after practice.

MacLeod, 1998). In our study, there was no associated change in the individual dimensions, as neither word reading nor color naming in the neutral condition was affected. In contrast, Dulaney and Rogers (1994) reported a "reading suppression" effect for young adults. However, they used 2240 practice trials compared with only 576 trials in our study. Perhaps longer training periods are required for the development of significant word reading inhibition. Thus, it remains unclear whether suppression or some other aspect of attentional efficiency may underlie our findings with adult Group A. In contrast, reading inhibition was evident in children practicing color naming of incongruent words, manifested as a relative increase in word reading latencies compared to color naming latencies (with neutral stimuli). Thus it seems that children in Group A learned to inhibit the verbal reading response, as predicted by the blocking hypothesis but not the automaticity hypothesis.

Attentional processes and executive control related to conflict detection and inhibition are clearly important in producing and modulating interference (MacLeod & MacDonald, 2000; Roelofs, 2003). It seems well established that interference is related to inhibition (Miyake & Friedman, 2012; van der Sluis, de Jong, & van der Leij, 2007), at least for good readers (Cox et al., 1997), and that individuals with attention deficits show increased interference (Lansbergen, Kenemans, & van Engeland, 2007). More generally, interference is used to assess cognitive control or executive function (van der Elst, van Boxtel, van Breukelen, & Jolles, 2006). We are not claiming that interference is primarily determined by reading ability. Our focus was on the relationship of interference with the individual underlying dimensions and not on the cognitive mechanism that might detect and resolve the conflict between them. Others have addressed the nature of this mechanism (e.g., Botvinick et al., 2001; Lovett, 2005; Roelofs, 2003), which remains of primary importance. Nevertheless, our results cannot be attributed to generic attentional learning because they were specific to one kind of practice.

Our findings suggest that the increased interference observed in childhood and aging may be due to different factors. Dulaney and Rogers (1994) proposed that old adults have difficulty developing new automatic responses, such as reading suppression. Our data indicate that this is not the case for children. Roelofs (2003) proposed that the U-shaped function of interference is due to the establishment of attentional control from childhood to adulthood that becomes less effective with aging. Our findings may be used to extend this proposal: It seems that attentional control in childhood is flexible (though relatively inefficient) whereas in aging it is not only less effective but also more rigid. In old adults executive and attentional factors might account for increased interference, whereas in childhood reading performance must also be taken into account. Thus, similar performance does not imply identical processing mechanisms (Dulaney & Rogers, 1994).

There was no evidence for reverse interference in our data even though reading performance is presumably not fully developed in children. This reinforces the notion of an inherent structural asymmetry between the two dimensions rather than an incidental asymmetry due to differential practice history. The emergence of reading suppression but not of reverse interference lends further supports to structural asymmetry: If asymmetry were due to practice then a new automatic response should have emerged in children practicing neutral color stimuli. Even if such an automatic response did develop, it was not evident in reading incongruent stimuli, as reading remained faster than color naming due to direct access to lexical form encoding.

In summary, we have tested a set of contrastive predictions arising from the automaticity and the blocking hypothesis regarding Stroop interference. The data showed that in children, where there is room for improvement in reading performance, word reading practice reduced Stroop interference. In adults, where maximum individual reading performance has been presumably reached, there was no such effect. Color naming practice did not affect interference in either population. These findings call for a reexamination of our conception of Stroop interference and the two performance dimensions underlying it.

References

Baayen, R. H. (2008). *Analyzing linguistic data*. Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

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 v. 0.999999-0. http://CRAN.R-project.org/package=lme4

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.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*, 332–361.

Comalli, P. E., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood and aging. *The Journal of Genetic Psychology*, *100*, 47–53.

Cox, C. S., Chee, E., Chase, G. A., Baumgardner, T. L., Schuerholz, L. J., Reader, M. J., & Denckla, M. B. (1997). Reading proficiency affects the construct validity of the Stroop test interference score. *Clinical Neuropsychology*, *11*, 105–110.

Davidson, D. J., Zacks, R. T., & Williams, C. C. (2003). Stroop interference, practice, and aging. *Aging, Neuropsychology, and Cognition*, *10*, 85–98.

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.

Durgin, F. H. (2000). The reverse Stroop effect. *Psychonomic Bulletin and Review*, *7*, 121–125.

Durgin, F. H. (2003). Translation and competition among internal representations in a reverse Stroop effect. *Perception & Psychophysics*, *65*, 367–378.

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.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116–124.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, *118*, 13–42.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446.

Kapoula, Z., Lê, T. T., Bonnet, A., Bourtoire, P., Demule, E., Fauvel, C., & Yang, Q. (2010). Poor performance in 15-year old dyslexic teenagers. *Experimental Brain Research*, *203*, 419–425.

Lansbergen, M. M., Kenemans, J. L., & van Engeland, H. (2007). Stroop interference and attention-deficit/hyperactivity disorder: A review and meta-analysis. *Neuropsychology*, *21*, 251–262.

Lovett, M. C. (2005). A strategy-based interpretation of Stroop. *Cognitive Science*, *29*, 493–524.

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 and Cognition*, *26*, 201–211.

MacLeod, C. M. (2005). The Stroop task in cognitive research. In A. Wenzel, & D. C. Rubin (Eds.), *Cognitive methods and their application to clinical research* (pp. 17–40). Washington, DC: Amer. Psychol. Assoc.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 126–135.

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, *4*, 383–391.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*, 8–14.

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.

Protopapas, A., Tzakosta, M., Chalamandaris, A., & Tsiakoulis, P. (2012). IPLR: An online resource for Greek word-level and sublexical information. *Language Resources and Evaluation*, *46*, 449–459.

R Core Team (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Found. Stat. Comput. http://www.R-project.org/

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review*, *110*, 88–125.

Sugg, M. J., & McDonald, J. E. (1994). Time course of inhibition in color-response and word-response versions of the Stroop task. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 647–675.

van der Elst, W., van Boxtel, M. P. J., van Breukelen, G. J. P., & Jolles, J. (2006). The Stroop color-word test: Influence of age, sex, and education; and normative data for a large sample across the adult age range. *Assessment*, *13*, 62–79.

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence*, *35*, 427–449.

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory and Cognition*, *13*, 304–319.