

Stroop interference depends also on the level of automaticity of the to-be-interfered process[☆]

Laurent Grégoire^{*,1}, Bénédicte Poulin-Charronnat, Pierre Perruchet

LEAD, CNRS UMR5022, Université Bourgogne Franche-Comté, Pôle AAFE, 11 Esplanade Erasme, 21000 Dijon, France

ARTICLE INFO

Keywords:
Stroop effect
Interference
Automatism
Practice
Musical expertise

ABSTRACT

The size of the Stroop effect is usually taken as dependent on the level of practice of the more automatized of two competing processes (e.g., reading in the standard Stroop task), possibly modulated in children by the age-dependent ability to inhibit nonrelevant information. However, this conclusion stems from experimental settings where the automaticity of the second process (e.g., color naming) is hard to assess and manipulate. The musical Stroop task, in which a note name is written inside a note on a staff, overcomes this limit. In the present experiment, children engaged in musical education were asked to read the written note names while ignoring the notes on the staff, or conversely, to name the notes while ignoring the written names. Both a Stroop-like effect and its reverse were observed, but, unexpectedly, the two effects did not evolve in parallel even though both musical and reading abilities improved during practice. Introducing the level of immunity to interference of the to-be-interfered process as a predictor of Stroop interference, in addition to the strength of the interfering process, appears as the best way to account for the interactive pattern.

1. Introduction

Most studies following the seminal paper by Stroop (1935) have investigated Stroop interference using the same two competing processes as in the original study, namely color naming and word reading. However, a number of variants have also been exploited. These variants were devised first to examine the generality of the effect over conditions and situations, but also, and maybe mainly, to explore theoretical questions that the properties of color naming and word reading made difficult or impossible to address. For instance, the picture-word task, which is the most widely used alternative to the color-word version, has been introduced because of “its greater flexibility in allowing many manipulations not possible with the restricted set of colors in the color-word task” (MacLeod, 1991, p. 167).

1.1. The musical Stroop paradigm

In the paradigm introduced by Grégoire, Perruchet, and Poulin-Charronnat (2013), the stimuli are composed of a note picture presented at several positions on a musical staff (see Fig. 1). The name of a note is printed inside the note picture. The written name is either congruent or incongruent with the position of the note in the staff. When asked to read the written note names while ignoring the positions of the note in the staff, participants having received musical education are slowed down when the written names are incongruent with the note positions. This Musical Stroop Effect (MSE) was interpreted as reflecting the automaticity of note naming in musicians, as the standard Stroop effect is interpreted as reflecting the automaticity of word reading. Note that word reading is still involved, but its status is inverted from the interfering dimension, as in classical Stroop-like paradigms, to the to-be-interfered dimension.²

[☆] This research was supported by grants from the Agence Nationale de la Recherche (Grant ANR-09-JCJC-0129) and from the Conseil Régional de Bourgogne (Grant No. 2012-9201AAO05S01353).

* Corresponding author.

E-mail addresses: lgregoire1@tamu.edu, laurent.gregoire@u-bourgogne.fr (L. Grégoire).

¹ Affective Attention Lab, Department of Psychology, Texas A&M University, College Station, TX 77840, United States of America.

² As in earlier papers exploiting the musical Stroop paradigm (e.g., Grégoire et al., 2015), the MSE is attributed here to interference, even though a part of the effect may arguably be due to facilitation in the congruent condition. In fact, teasing apart interference and facilitation would imply the use of a methodologically unquestionable neutral condition, which has proven to be difficult to achieve in the musical Stroop paradigm (Grégoire et al., 2013). Identifying Stroop effect and Stroop interference is an oversimplification. However, it seems acceptable in a first step insofar as interference is always reported as much larger than facilitation. Presumably due to task conflict, sometimes RTs in congruent condition are even slower than RTs in neutral condition (the so-called reverse facilitation effect, e.g., Goldfarb & Henik, 2007), especially in young children (e.g., Ben-Shalom, Berger, & Henik, 2013).

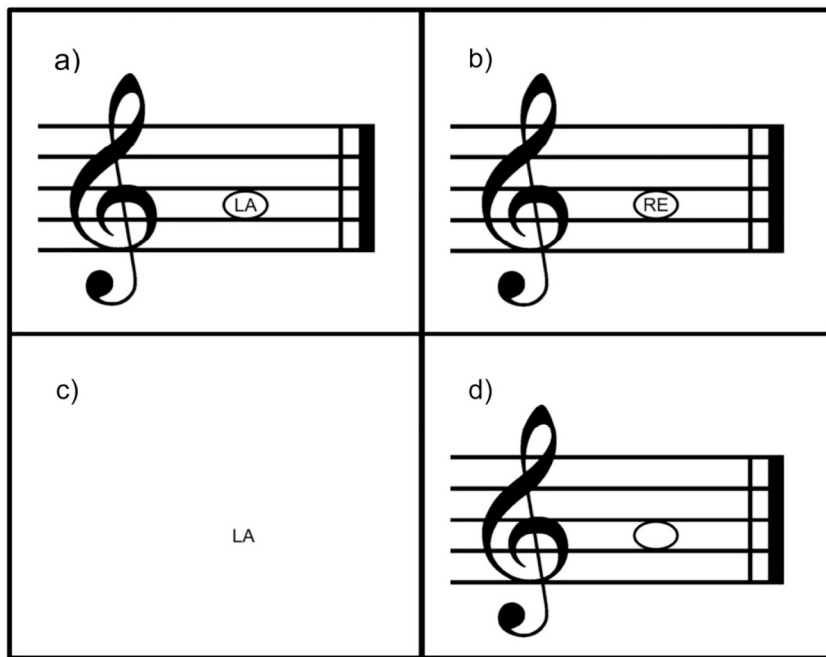


Fig. 1. Examples of the different conditions used in the experiment: a) congruent condition; b) incongruent condition; c) reading-ability test; and d) note-naming ability test. Note that in the musical French notation (and several other countries such as Italy and Spain), note names are DO, RE, MI, FA, SOL, LA, SI, instead of the first letters of the alphabet.

Up to now, this new paradigm has been exploited to explore two main issues, which are not easily dealt with any other Stroop-like paradigms. The first one is the change in Stroop interference with the amount of practice. Of course, a number of earlier studies have investigated the changes in the Stroop effect as a function of reading skills, but interpreting their results is problematic because literacy training is generally confounded with age. Typically, these changes are described as following an inverted U-shape function, whereby an initial increase is followed by a significant decrease, with a peak being located after 2 or 3 years of reading practice (Armengol, 2002; Dash & Dash, 1982; Peru, Faccioli, & Tassinari, 2006; Rand, Wapner, Werner, & McFarland, 1963; Schadler & Thissen, 1981; Schiller, 1966). In a few studies, the initial increase of interference was even absent, with only the downward component of the curve being observed (Comalli Jr., Wapner, & Werner, 1962; Rosinski, Golinkoff, & Kukish, 1975).

These data are challenging, because if Stroop interference in the conventional setting reflects the automaticity of reading, and if the level of automaticity is a growing function of the amount of practice, then the capacity of reading to interfere with another activity should appear gradually, and should increase as reading skills improve. A common interpretation of the paradoxical data stems from the fact that at least a part of the decreasing component of the curve may be due to general factors evolving during childhood. Comalli Jr. et al. (1962) were the first to suggest that the amount of interference could be a positive function of the amount of practice but that this effect would be overshadowed by the age-related variations in the ability to inhibit non-relevant information. Age-related variations in cognitive control would explain the decrease in Stroop interference observed from childhood to adulthood, because it is commonly acknowledged that this ability grows during the relevant period of time (e.g., Bedard et al., 2002; Klenberg, Korkman, & Lahti-Nuutila, 2001). This interpretation received support from a few studies, in which age was held constant by separating fast and slow readers in adults (Martin, 1978) or in children within a given academic level (Stanovich, Cunningham, & West, 1981). As a rule, fast readers exhibited greater interference.

The musical Stroop paradigm was exploited to test the hypothesis that the amount of interference is a positive function of practice when age is no longer a confounding factor. Admittedly, most children begin studying music at the same age, around the age of 5 or 6, hence raising the same potential confound between age and level of practice as for

reading. However, the association between age and practice is much looser, because musical training may also begin sooner or later in children, and even during adulthood. As a consequence, appropriate selection of participants makes it possible to decouple the effects of age and practice. Using the musical Stroop paradigm in children of similar age but with levels of musical training varying from one to five years of practice, Grégoire, Perruchet, and Poulin-Charronnat (2015) observed a positive relation between the amount of interference and the extent of musical training. These results lead to the important conclusion that the interfering power of a given process grows as the automatization of this process progresses, provided that the improvement in cognitive control abilities due to age does not overshadow this effect.

A second issue that the musical Stroop paradigm has concurred to enlighten is related to whether Stroop interference may be bidirectional. In the standard color-word version of the task, and in most Stroop-like tasks, Stroop interference is essentially unidirectional, or at least asymmetrical. In the standard version of the task, for instance, color naming is hampered by the reading of an incongruent color name, but reading a color word is virtually unaffected when the word is written in an incongruent color ink. A reverse effect has been reported, but only when word reading was strongly degraded (Dunbar & MacLeod, 1984), or when verbal responding was replaced by motor responding (e.g., Blais & Besner, 2006, 2007; Durgin, 2003; Melara & Mounds, 1993). Other studies using reading but another competing dimension than color revealed a reverse Stroop effect. In Akiva-Kabiri and Henik (2012), for instance, musician participants were asked to read the name of notes or to name notes while hearing a tone that either corresponded to the note or not. Absolute pitch possessors showed a significant congruity effect, indicating that pitch identification was impossible to suppress. However, when absolute pitch possessors were asked to identify auditory tones while ignoring the written note name or the note picture, they were unaffected by the irrelevant visual stimuli. As a consequence, although they observed a reverse Stroop effect in the sense that reading was the to-be-interfered dimension instead of being the interfering dimension, their results do not challenge the unidirectionality of interference: If one effect is present, its reverse is absent and vice versa (see also Palef & Olson, 1975).

Grégoire, Perruchet, and Poulin-Charronnat (2014) provided unambiguous evidence for bidirectional interference in the very same experimental setting. The task given to musicians for obtaining an MSE

is a word-reading task, with note naming acting as a potentially interfering process when the note name written inside the note is incongruent with the location of the note on the staff. However, with the very same materials, it is also possible to ask the musicians to name the notes as a function of their position, thus reversing the function of word reading and note naming. Naming a note should take longer when the note name written inside the note is incongruent with the note location on the staff than when it is congruent. With adult musicians, Grégoire et al. observed exactly this effect, which they named the RMSE, for “Reverse MSE”. In fact, but may be unsurprisingly, on average, the RMSE, due to the automaticity of reading, was stronger than the MSE, due to the automaticity of note naming in musicians. Grégoire et al. noted that this strong evidence for bidirectional interference run against some theoretical approaches, especially the “horse race” model (e.g., Morton & Chambers, 1973) and the Cohen, Dunbar, and McClelland (1990) connectionist model.

Note that observing bidirectional interference is consistent with the results from another Stroop-like paradigm in which word reading is no longer involved, by contrast with the studies reviewed so far. In Henik and Tzelgov (1982), digits were presented visually in fonts of different size, and participants carried out either a physical size judgment while ignoring the numerical values, or a numerical size judgment while ignoring the physical size. A numerical Stroop effect was found in both tasks (see also Arend & Henik, 2015; Glikzman, Itamar, Leibovich-Raveh, Melman, & Henik, 2016; MacLeod & Dunbar, 1988). However, there is an important difference between the musical and the numerical Stroop effects. Digit reading certainly improves with school practice as word reading does, but it is hard to conceive how size processing could be measured, manipulated, and even thought of as a practice-dependent ability. In fact, it could even be argued that what is learned during the school years is rather that physical size is generally irrelevant when reading words or digits, and should be discarded. By contrast, the musical Stroop paradigm allows the investigation of mutual interference between two automatisms that should grow concomitantly.

1.2. The present study

As just recalled, the musical Stroop paradigm was exploited to investigate the evolution of MSE with practice on the one hand (Grégoire et al., 2015), and the presence of both a MSE and a RMSE as a function of task instructions on the other hand (Grégoire et al., 2014). The present study further examines the evolution of interference with practice, but takes profit of the opportunity of observing this evolution in parallel on the MSE and the RMSE. The primary interest of collecting both the MSE and the RMSE is to provide considerable constraints over the whole system, because it would be nonsensical to endorse an interpretation on the ground that it is well-suited for an effect if it turns out to be inconsistent with the other effect.

As in Grégoire et al. (2015), in the following experiment, children coming from the first five years of musical training served as subjects. But contrary to Grégoire et al. (2015), in which the age of participants was maintained constant across the five years of musical training, in the following experiment, the age of children increased, on the average, by one year when going from a given level of musical training to the next one, and likewise for their modal grade level. Increasing the age of children results in increasing their practice in word reading. As a consequence, in the following experiment, both word-reading and note-naming abilities should evolve in parallel. All children had to carry out two tasks, devised to record the MSE and the RMSE respectively. In the word-reading task (MSE), children were required to read written note names, while ignoring the positions of the note in which the note names were printed. In the note-naming task (RMSE), the same children were asked to name notes, while ignoring the written note names printed inside. Furthermore, their word-reading and note-naming abilities were measured in independent tests. So, in keeping with a widespread tradition (e.g., Cohen et al., 1990), the strength of the two competing

processes was assessed through both the amount of practice and the speed of processing.³

Two different predictions about the expected effects have been discussed above. The first prediction stems from the common view that increasing the strength of a process through prolonged practice increases the degree of irrepressibility of this process. In this view, the MSE should increase as a proportion of the strength of note naming, and the RMSE should increase as a proportion of the strength of word reading.

A second possibility is that the two effects decrease, whatever the strength of the respective processes, due to the fact that the amount of training is confounded with age. The general enhancement in cognitive control of children while they grow in age would be responsible for this decrease. As mentioned above, this interpretation has been put forward long ago to explain the observed decrease in size of the color-word Stroop effect during childhood (e.g., Comalli Jr. et al., 1962), but the idea that age-related changes in cognitive control modulate the Stroop effect is still ongoing (e.g., La Heij & Boelens, 2011).

The two accounts above, although prevalent, are not the only possible ones, however. We are aware of two other, not yet evoked, predictions about the effect of practice on the amount of interference, which have received supporting evidence in the Stroop literature. First, practice could improve control irrespective of age variations. Logan (1985) wrote that “skilled performers are usually able to control their performance better than unskilled performers, even though their performance is likely to be more automatic,” (p. 379). Evidence for a better control has been observed by Logan (1982), who demonstrated that skilled typists were able to inhibit high-speed typing when detecting an error or an overt signal to stop. This idea was applied to the Stroop effect by Tzelgov, Henik, and Leiser (1990). By testing bilinguals, Tzelgov et al. observed that Stroop interference is controllable, and that language proficiency is a precondition for such a control. Participants were able to reduce Stroop interference as a function of their expectancies in their native language but not in their second language. It is thus possible that the decrease in Stroop interference observed from childhood to adulthood reflects a genuine property of automatisms, whereby the possibility of cognitive control would increase with practice. In this view, the MSE should decrease as a proportion of the strength of note naming, and likewise, the RMSE should decrease as a proportion of the strength of word reading.

Finally, the amount of interference could be also modulated through a change in the strength of the competing dimension. In the Cohen et al.'s (1990) model, increasing the strength of a Process A both increases its ability to produce interference on a Process B (as pointed out above) and, crucially, reduces its susceptibility to the interference coming from Process B. Because the strength of both word reading and note naming should increase with practice, the MSE should decrease during training as a proportion of the strength of word reading, and conversely, the RMSE should decrease as a proportion of the strength of note naming.

For the sake of concision, the four sources of influence envisioned above, which are not exclusive one of the other, will be coined, respectively, as *Interference*, *Age-related control*, *Practice-related control*, and *Resistance to interference*. The present study is aimed at determining which of these four processes, or combination of processes, provides the best account for our results. In theory, the action of these processes

³ Assessing the strength or the level of automaticity of a given process through speed measures was discussed at length in Grégoire et al. (2014). It is worth stressing that exploiting speed does not amount to endorsing the so-called horse-race model of the Stroop effect. The claim that the direction of interference depends on which of the two competing processes is completed first has been clearly rejected, among others by Glaser and Glaser (1982) and Dunbar and Macleod (1984) studies. However, it is commonly admitted that for a given task, speed measures allow the comparison of the strength of the underlying processes at different stages of practice, or between different participants.

Table 1
Predictions on the evolution of the MSE and RMSE with practice, assuming different sources of influence.

Sources of influence	MSE (task = word reading, WR)	RMSE (task = note naming, NN)
Interference	<i>increases</i> as a function of the strength of NN	<i>increases</i> as a function of the strength of WR
Age-related control	<i>decreases</i> as a function of age	<i>decreases</i> as a function of age
Practice-related control	<i>decreases</i> as a function of the strength of NN	<i>decreases</i> as a function of the strength of WR
Resistance to interference	<i>decreases</i> as a function of the strength of WR	<i>decreases</i> as a function of the strength of NN

leads to different predictions, as summarized in Table 1. However, dissociating some of these predictions raises a problem. Assuming for instance that interference on Process A coming from Process B decreases across the years of training, how could we know whether this decrement is due to the increased *Resistance to interference* of Process A or to a better *Practice-related control* of Process B, given that both Processes A and B are practiced? If the strengths of Processes A and B evolve in exactly the same way, distinguishing between these accounts would be impossible. Fortunately, in our study, the strength of note naming and the strength of word reading should not start from the same level, and should not increase to the same rate. Indeed, children will be examined at a time where they are engaged in the first steps of musical instruction, whereas they began to read a few years earlier. As a consequence, reading abilities should be more developed than musical abilities for the youngest participants. In keeping with the ubiquitous power law of learning, this initial lag should trigger a sharper improvement in note naming than in word reading across the years of practice. To anticipate, these prerequisites will be clearly met in our study. As a consequence, to consider again the example above, distinguishing between interpretations becomes possible. Indeed, the observed decrease in interference will be attributed to the increased *Resistance to interference* of Process A, or alternatively to a better *Practice-related control* of Process B, according to whether the observed decrease in interference fits better with the changes observed in the ability A or B, respectively.

2. Method

2.1. Participants

One hundred children from the *Conservatoire à Rayonnement Régional* of Dijon took part in the experiment. The children were divided into five groups according to their number of years of musical training (from 1 to 5). As expected, the age of children increased, on the average, by one year when going from a given level to the next one, with age ranging approximately from eight to 12 from the first to the fifth level of musical education. Likewise, the modal academic grade co-varied with the musical level, with grade ranging from 2 to 6 from the first to the fifth level of musical education (see Table 2). All participants were French native speakers and reported normal or corrected-to-normal vision. A parental consent was obtained for each child.

2.2. Materials

The stimuli consisted of a treble staff with a note picture, which could appear on each of the seven possible positions going from C4 to B4. The name of a note was written inside the note picture. For the congruent condition, the note name was congruent with the note position on the staff (Fig. 1a), whereas in the incongruent condition, note name and position were incongruent, with the name written inside the note picture being one of the six other possible note names (e.g., when the note was DO, the written name was LA, SI, RE, MI, FA, or SOL; Fig. 1b).

The stimuli used in additional tests of word-reading and note-naming abilities were designed to be closer from those involved in everyday situations. They are represented in Fig. 1c and d, respectively.

To prevent the iconic memory of the staff to influence the processing of the following note, the stimuli were randomly displayed at one of four possible positions without immediate repetition at the same location. The four positions were defined as the center of (invisible) rectangles resulting from the exhaustive partitioning of the screen into four quadrants of equal size. Stimuli were printed in black over a white background on a computer screen. Note names appeared in standard uppercase font 14. The treble staff was 7.7 cm wide by 5.1 cm high.

2.3. Procedure

The children had to perform a word-reading task and a note-naming task in succession. The order of presentation of the two tasks was counterbalanced across children. In the word-reading task, the children had to read aloud the printed word while ignoring the note picture. In the note-naming task, the children were asked to name the note while ignoring the word written inside.

For each task, there were two mixed conditions: congruent (Fig. 1a) and incongruent (Fig. 1b). For each condition, the stimuli appeared six times on each of the 7 locations, resulting in 42 trials per condition, and 84 trials (42×2) for each task. On each trial, a fixation cross was displayed for 1 s at the center of the screen before the apparition of the stimulus, which stayed on the screen until participant's response. The interval between the response and the next trial was 1 s. The trials were pseudo-randomly ordered for each participant, excluding immediate repetitions of note locations or note names. They were displayed as four blocks of 21 trials each with a self-paced break between blocks.

The experimental session was immediately followed by two additional tests, which were run in counterbalanced order. One test was a reading-ability test, in which the children had to read note names (Fig. 1c). The other test was a note-naming ability test, in which the children had to name notes (Fig. 1d). Each test included 21 trials. The trials were pseudo-randomly ordered for each participant, excluding immediate repetitions of words or notes.

Whatever the tasks, participants were encouraged to respond as fast and as accurately as possible throughout the session. The response times (RTs) were recorded by a voice key. During the session, the experimenter noted error responses and voice-key dysfunctions. After the experiment, the children filled out a questionnaire about their musical training.

3. Results

3.1. Ability tests

Voice-key dysfunctions led to exclude 3.83% of the data. Children's errors are shown in Table 2. An ANOVA performed with Task (word reading, note-naming) as a within-subject variable and Years of Musical Training (1, 2, 3, 4, 5) as a between-subject variable showed a significant main effect of task, with more errors in the note-naming task, $M = 2.46\%$, $SD = 5.03$, compared with the word-reading task, $M = 0.24\%$, $SD = 1.25$, $F(1, 95) = 18.83$, $p < .001$, $\eta_p^2 = 0.165$. There was no main effect of years of musical training, $F(4, 95) = 1.83$, $p = .130$, and no Task \times Years of Musical Training interaction, $F(4, 95) = 1.81$, $p = .133$.

RTs for correct responses beyond three standard deviations of the

Table 2

Participant characteristics and error percentages for the ability tests, word-reading task, and note-naming task as a function of musical training (for Age and error percentages, numbers in parenthesis are Standard Errors of the Mean, SEM).

	Years of musical training				
	1 year	2 years	3 years	4 years	5 years
Grade 2	15	4	–	–	–
Grade 3	4	12	1	–	–
Grade 4	1	4	10	2	–
Grade 5	–	–	5	16	6
Grade 6 and >	–	–	5	2	13
Age (years)	8.20 (0.70)	8.99 (0.76)	10.94 (1.81)	11.01 (0.67)	12.16 (1.38)
N total (females)	20 (10)	20 (9)	21 (14)	20 (13)	19 (13)
Word-reading ability test	0.25 (1.12)	0.71 (2.33)	0 (0)	0.24 (1.17)	0 (0)
Note-naming ability test	5.14 (8.68)	1.74 (3.34)	1.60 (2.76)	1.73 (3.29)	2.11 (4.19)
Word-reading congruent	0.12 (0.55)	0.12 (0.53)	0.47 (1.26)	0.24 (0.74)	0.13 (0.57)
Word-reading incongruent	0.47 (1.23)	0.77 (2.37)	0.71 (1.38)	1.35 (2.46)	1.48 (2.43)
Note-naming congruent	3.14 (4.46)	1.97 (3.74)	1.60 (4.15)	0.24 (0.75)	1.06 (3.52)
Note-naming incongruent	9.23 (9.11)	10.12 (6.98)	7.55 (6.54)	6.11 (5.09)	7.99 (7.86)

mean (0.85%) were removed. The remaining data are shown in Fig. 2, top panel. An ANOVA on children's correct RTs performed with Task (word reading, note-naming) as a within-subject variable and Years of Musical Training (1, 2, 3, 4, 5) as a between-subject variable showed a significant main effect of task, with shorter RTs for the word-reading task compared with the note-naming task, $F(1, 95) = 99.64, p < .001, \eta_p^2 = 0.512$, and a significant main effect of years of musical training, with faster RTs as the number of years of musical training increased, $F(4, 95) = 28.40, p < .001, \eta_p^2 = 0.545$.

These two main effects were qualified by a significant Task \times Years of Musical Training interaction, $F(4, 95) = 20.32, p < .001, \eta_p^2 = 0.461$ (Fig. 2, top panel). For the word-reading ability test, there was a significant main effect of years of musical training, $F(4, 95) = 10.49, p < .001, \eta_p^2 = 0.306$, with a significant decreasing linear trend, $F(1, 95) = 32.88, p < .001, \eta_p^2 = 0.257$. A similar pattern of results was observed for the note-naming ability test, with a significant main effect of years of musical training, $F(4, 95) = 25.56, p < .001, \eta_p^2 = 0.518$, and a significant decreasing linear trend, $F(1, 95) = 72.22, p < .001, \eta_p^2 = 0.432$. However, as anticipated, the decreasing linear trend was significantly greater for the note-naming ability test than for the word-reading ability test, $F(1, 95) = 56.22, p < .001, \eta_p^2 = 0.372$.

Note that there was no evidence for a speed-accuracy trade-off whatever the test. Indeed, the correlation between errors and RTs was either null (word-reading test, $r(98) = -0.016, p = .876$) or positive (note-naming test, $r(98) = 0.263, p = .008$), whereas a trade-off would have elicited negative correlations.

3.2. Stroop tasks: overall ANOVAs

Voice-key dysfunctions led to exclude 3.46% of the data for the MSE and 5.44% of the data for the RMSE. The percentage of children's errors is shown in Table 2 for each condition. An ANOVA performed on these data with Task (word reading, note naming) and Congruity (congruent, incongruent) as within-subject variables, and Years of Musical Training (1, 2, 3, 4, 5) as a between-subject variable, revealed a significant main effect of task, $F(1, 95) = 91.30, p < .001, \eta_p^2 = 0.490$, with more errors in the note-naming task ($M = 4.90\%, SD = 6.59$) than in the word-reading task ($M = 0.59\%, SD = 1.58$). Importantly, the main effect of congruity was also significant, $F(1, 95) = 98.67, p < .001, \eta_p^2 = 0.509$, with more errors in the incongruent condition ($M = 4.57\%, SD = 6.41$) than in the congruent condition ($M = 0.92\%, SD = 2.72$). However, this effect was qualified by a Task \times Congruity interaction, $F(1, 95) = 66.83, p < .001, \eta_p^2 = 0.413$. Additional ANOVAs were performed to examine whether a congruity effect was present for each task. For word reading, there was a significant main effect of congruity, $F(1, 95) = 13.52, p < .001, \eta_p^2 = 0.125$, with more errors in the

incongruent condition ($M = 0.95\% SD = 2.03$) than in the congruent condition ($M = 0.22\%, SD = 0.79$). Similar data were obtained for note naming. There was a significant effect of congruity, $F(1, 95) = 89.02, p < .001, \eta_p^2 = 0.484$, with more errors in the incongruent condition ($M = 8.20\%, SD = 7.21$) than in the congruent condition ($M = 1.61, SD = 3.64$). Thus, both the MSE and the RMSE were reliable, the Task \times Congruity interaction being due to the fact that the RMSE ($M = 6.59\%, SD = 6.90$) was much larger than the MSE ($M = 0.73\%, SD = 2.02$).

There was no main effect of years of musical training, $F(4, 95) = 0.96, p = .430$, and no interaction between years of musical training and congruity, $F(4, 95) = 0.49, p = .745$, suggesting that when the MSE and the RMSE are assessed through error rates, there is no change with training. Visual inspection of Table 2 suggests that this lack of effect could be due to the fact that the MSE tended to increase throughout training, whereas the RMSE tended to decrease. For instance, comparing the first two years of training to the last two years, the MSE increased from 0.50% to 1.23%, whereas the RMSE decreased from 7.12% to 6.40%. These trends are of potential interest because, to anticipate, the very same pattern will be observed on RT. However, regarding errors, there was no statistical support for this analysis, as the Task \times Congruity \times Years of Musical Training interaction was not significant, $F(4, 95) = 0.39, p = .814$.

The same overall analysis was performed on RTs. Correct responses beyond three standard deviations of the mean RT were removed. Given that RTs substantially differed between tasks, this exclusion criterion was applied separately for each task. This leads to remove 1.00% of the RTs for word reading and 1.46% of the RTs for note naming. The remaining data are shown in Fig. 2 for each condition. An ANOVA performed on these data with Task (word reading, note naming) and Congruity (congruent, incongruent) as within-subject variables, and Years of Musical Training (1, 2, 3, 4, 5) as a between-subject variable, revealed a main effect of task, $F(1, 95) = 215.09, p < .001, \eta_p^2 = 0.694$, with shorter RTs in the word-reading task compared with the note-naming task, a main effect of congruity, $F(1, 95) = 366.24, p < .001, \eta_p^2 = 0.794$, with shorter RTs in the congruent condition compared with the incongruent condition, and finally a main effect of years of training, $F(4, 95) = 27.01, p < .001, \eta_p^2 = 0.532$, with RTs decreasing monotonously across years. However, all the two-way interactions were also significant: Task \times Congruity, $F(1, 95) = 209.75, p < .001, \eta_p^2 = 0.688$, Task \times Years of Musical Training, $F(4, 95) = 22.40, p < .001, \eta_p^2 = 0.485$, and Congruity \times Years of Musical Training, $F(4, 95) = 12.87, p < .001, \eta_p^2 = 0.351$. Moreover, these interactions were qualified by a significant Task \times Congruity \times Years of Musical Training interaction, $F(4, 95) = 13.37, p < .001, \eta_p^2 = 0.360$. Given this pervasive interactive pattern, separate ANOVAs were performed for each task.

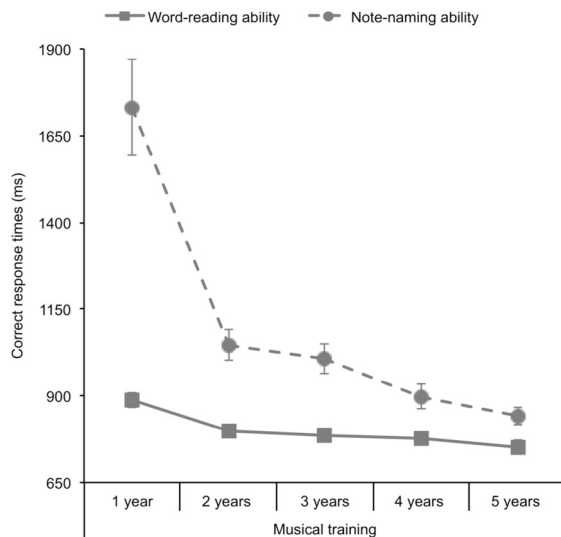
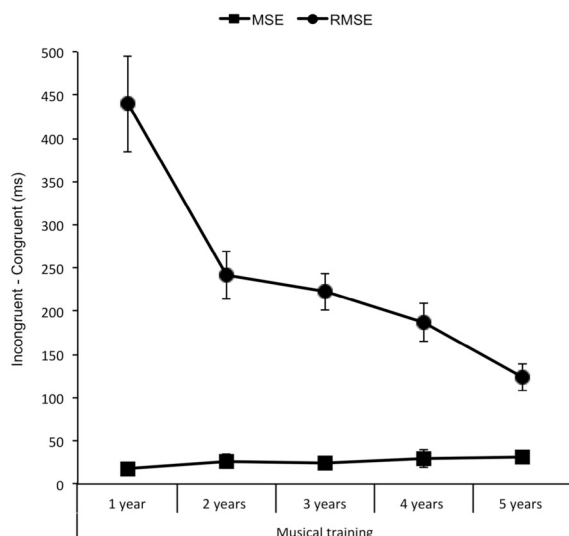
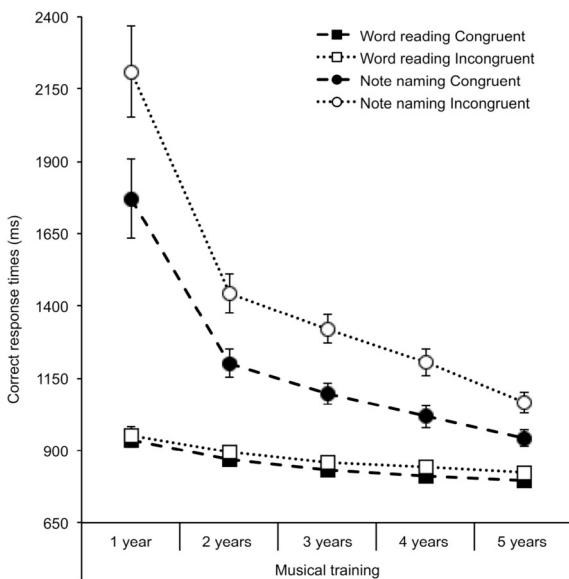


Fig. 2. Correct response times for the word-reading ability test and the note-naming ability test as a function of Years of Musical Training (top panel); correct response times for the word-reading task and for the note-naming task as a function of Congruity and Years of Musical Training (middle panel). The amplitude of the MSE and the RMSE (i.e., incongruent minus congruent) is also plotted (bottom panel). Error bars indicate standard errors.



As mentioned above, a second estimation of the strength of an automatic pathway is given by the speed of processing. Strength was assessed here as the opposite of the mean RT (i.e., $RT^* - 1$) in the ability tests, in keeping with the widely accepted postulate that lower RTs are indicative of higher strength and conversely (the resulting numbers were negative, but this was inconsequential for the subsequent calculation of correlations). In the subsections below, specific ANOVAs are completed by correlational analyses, in which correlations between strength and the amount of interference (assessed through both errors and RTs) were computed across the whole sample of children ($N = 100$).

3.3. MSE

For word reading, there was a significant main effect of congruity, $F(1, 95) = 60.06, p < .001, \eta_p^2 = 0.387$, with longer RTs in the incongruent than in the congruent condition attesting for a MSE, and a significant main effect of years of musical training, $F(4, 95) = 8.68, p < .001, \eta_p^2 = 0.268$, with decreasing RTs with years of musical training. The Congruity \times Years of Musical Training interaction was not significant, $F(4, 95) = 0.43, p = .786$. Further t -tests showed that RTs were longer in the incongruent condition than in the congruent condition for all groups, $t(19) = 2.54, p = .020, d = 0.17, t(19) = 3.40, p = .003, d = 0.41, t(20) = 5.23, p < .001, d = 0.32, t(19) = 2.88, p = .010, d = 0.52$, and $t(18) = 4.72, p < .001, d = 0.43$, for one, two, three, four, and five years of musical training, respectively.

The correlations between the size of the MSE and the strength of note naming were positive, suggesting that interference increased with the strength of the interfering process. When the amount of interference was computed from error rates, the correlation did not reach significance, $r(98) = 0.166, p = .099$ (two-tailed, as all the p -values reported in this paper). However, statistical significance was ensured when the correlation involved RTs, $r(98) = 0.228, p = .022$.

Thus, there is some discrepancy according to whether strength was assessed as the number of years of musical practice, or from the musical ability test. It is worth noting that this discrepancy is limited. On the one hand, although the effect of years of practice was significant neither for error rates nor for RTs, the MSE numerically increased for the two measures. From one to five years of musical training, the MSE measured through error rates increased by a factor of 3.8, and the differences in RTs increased from 18.16 ms to 30.67 ms. On the other hand, the correlations with the score of ability were moderate in size. A part of the explanation for the nonsignificant results from the ANOVA is that the number of years of training could be a noisy measure of automatization, especially for musical instruction, which may give rise to very different amount of home practice between children. An additional problem is that, depending on the instrument they played, not all children may be trained to the same extent on the treble clef, which was used in the measure of interference. For instance, the choice of contrabass or cello implies to shift for bass clef for instrumental practice. In this regard, the score in the ability test could provide a more reliable information, because the very same part of the staff was used in this test and in the measures of interference. To avoid the recollection of the whole pattern of results each time the MSE is evoked, we refer hereafter to the “numerical increase” of the MSE with increasing strength (keeping in mind, behind this conservative terminology, that statistical evidence is mixed, but not absent).

3.4. RMSE

The ANOVA on note naming scores gave a significant main effect of congruity, $F(1, 95) = 295.68, p < .001, \eta_p^2 = 0.757$, with longer RTs for the incongruent than for the congruent condition attesting to a RMSE, and a significant main effect of years of musical training, $F(4, 95) = 26.04, p < .001, \eta_p^2 = 0.523$, with shorter RTs as the number of years of musical training increased. Unlike the analysis focusing on the MSE, these two main effects were qualified by a significant Congruity \times Years of Musical Training interaction, $F(4, 95) = 13.85, p < .001, \eta_p^2 = 0.368$. Crucially, the difference between incongruent and congruent conditions decreased from one to five years of musical training, following a significant linear trend, $F(1, 95) = 46.03, p < .001, \eta_p^2 = 0.326$.

Using RTs in the word-reading ability test as a measure of strength confirms the effect. The correlations between the size of the RMSE and the strength of word reading were negative, indicating that, in contrast with the relation observed for the MSE, interference numerically *decreased* when the strength of the interfering process increased. Using error rates resulted in a negligible value, $r(98) = -0.061, p = .289$, but with RTs as a measure of interference, the correlation revealed a medium to large effect, $r(98) = -0.402, p < .001$. Interestingly, this correlation and the corresponding correlation between the size of the MSE and the strength of note naming, were significantly different ($r = -0.402$ vs. $r = 0.228, p < .001$, test of significance performed with the “cocor.dep.groups.nonoverlap” function of the R package ‘cocor’, assessing the difference between two correlations of dependent groups with no variable in common). This difference echoes the three-way interaction reported above in the overall ANOVAs, and concurs to indicate that the upward change of the MSE across the years of musical training and the downward change of the RMSE significantly differed (see the curves at the bottom of Fig. 2).

4. Discussion

Several results from the present experiment replicated earlier data. The RMSE was substantially larger, on the average, than the MSE. This difference is consistent with the results of Grégoire et al. (2014), and more generally with the asymmetry commonly observed since Stroop (1935): When reading is involved as one of the two competing dimensions, the interference due to reading is much stronger than the interference on reading due to the other dimension, which is often altogether absent. However, both the MSE and the RMSE were reliable, replicating the observation of bidirectional effects in Grégoire et al. (2014). As discussed at length in Grégoire et al., this finding is worth noting, because getting both a Stroop-like effect and its reverse without changing the respective saliency of the two dimensions is quite infrequent in the Stroop literature, at least when reading is involved as one of the two competing processes. Finally, as in Grégoire et al. (2015), the MSE tended to increase with the strength of note naming. When strength was assessed as the speed of processing in the note-naming ability test, the correlation with the MSE was positive and significant, confirming the Grégoire et al.'s (2015) conclusion that interference increases with the level of automatization of the interfering dimension. However, when the strength of note naming was assessed as the number of years of musical education, the observed increments in both errors and RTs were not significant, while significant results for RTs were reported in Grégoire et al. (2015). A question of interest is whether this difference attests to a genuine flattening of the curve of the MSE in the present experiment, in regard of our prior study. In Grégoire et al. (2015), the MSE increased from 6 to 20 ms over three years of musical practice in Experiment 1, and from -10 to 19 ms over five years of musical practice in Experiment 2. Recall that the corresponding values in the present report were 18 and 31 ms, which fall in the same order of magnitude as in the earlier experiments. The difference in the level of significance is certainly due to that fact that the ANOVAs in

Grégoire et al. included a group of nonmusician children who, unsurprisingly, performed at chance. In the present experiment, the increase of the MSE was assessed by taking as starting point the above-chance performance of musician children in their first year of training.

All these successful replications are noteworthy, at a time where the issue of replicability in cognitive psychology and other experimental sciences has become a pressing object of concerns (e.g., Baker, 2016). However, the experiment was mainly devised to explore a new issue, namely the joint evolution of the MSE and the RMSE when the two competing processes engaged in the tasks became more and more automatized. The main new result was that the RMSE exhibited a striking decrease when going from the first to the fifth year of musical education. Considered jointly with the numerical increase of the MSE, the final picture is a clear-cut interactive pattern. In addition, the level of automaticity of the two competing processes, as assessed from the ability tests, improved during these five years, although, as expected, improvement was steeper for note naming than word reading.

This pattern of results allows a first, unquestionable conclusion: Taken in isolation, none of the four possible processes listed in the Introduction turns out to account for the data. The reason is straightforward: Whatever the involved process, both effects should increase or both effects should decrease. Considering *Interference* for instance, the numerical increase of the MSE is consistent with the idea that interference is a positive function of practice (the MSE, in which note naming is the interfering dimension, correlated positively with the strength of note-naming), but the decrease of the RMSE is incompatible with this hypothesis (the RMSE, in which word reading is the interfering dimension, correlated negatively with the strength of word reading).

In the remaining of the Discussion, we examine how introducing each of the other postulated processes in addition to *Interference* might improve the predictions made with *Interference* alone. For instance, the first account to be considered, namely *Age-related control*, integrates both *Interference* and *Age-related control*. The need for adding *Interference* to the other processes stems from the fact that *Interference* is the only source of influence that generates a Stroop-like effect in the first place (the other three sources of influence are assumed to act against this effect).

4.1. Adding age-related control

Recall that a widely shared view is that age-related variations in cognitive control may explain the decrease in color-word Stroop interference observed from 7 to 9-year-old children to adulthood, despite improved reading abilities. Grégoire et al. (2015) brought out some support to this view, by showing that when age was held constant, Stroop interference did not decline. Instead, the MSE slightly *increased* with musical practice. Given that musical students in the present study were tested while they were between 8 and 12-year-old, the improvement in age-related cognitive control should trigger a decrease of both the MSE and the RMSE. A decrease of the RMSE was actually observed, but there was again a slight increase of the MSE. Can this increase be conciliated with the growing influence of age-linked inhibitory factors?

One could argue that the same inhibitory processes are involved to reduce interference for the RMSE and the MSE, but that the positive effect of practice would be much stronger for the MSE. In other words, the hypothesis would be that without the downward effects of age-linked factors, the increase of interference with years of practice would be much larger for the MSE than for the RMSE. This hypothesis would be consistent with the fact that RTs in the musical-ability test exhibit a significantly steeper slope than RTs in the reading-ability test. However, this hypothesis strongly conflicts with the data reported in Grégoire et al. (2015). Indeed, as analyzed above, even though age was held constant (or at least varied within a very restricted range), the increase of the MSE was in the same order of magnitude as in the present experiment, making it implausible that age-linked factors may have

played a strong influence on the MSE in this experiment. Thus, assuming that the substantial decrease of the RMSE observed in our experiment is due to the growing control of children over their responses appears hard to conciliate with the apparent insensitivity of the MSE to the same factor.

4.2. Adding Practice-related control

As above, the evolution of Stroop interference would be the end-result of two opposite effects: A positive effect due to the fact that the interfering power of a process increases as this process becomes more and more automatized, and a negative effect acting to restrain this interference. However, rather than assuming that the negative effect is due to an ability to inhibit nonrelevant information thought of as growing with age during childhood, the *assumption is now* that interference is toned-down by the increased control due to extensive practice.

A better control on word reading with increased practice should lead to a decrease of the RMSE, which was empirically confirmed. However, the same account does not work for note naming. A better control on note naming with increased practice would lead to a decrease of the MSE, which was not observed. This seemingly rule out the hypothesis that the better control of intensively practiced skills, suggested by Logan (1985) and Tzelgov et al. (1990), may help to account for the present results. To make our pattern of results consistent with this framework, one must suppose that the better control applies only (or mainly) to word reading, and not to note naming. Are there some independent reasons that could make this hypothesis more than a purely ad-hoc speculation?

In fact, the hypothesis according to which practice improves control only for word reading and not for note naming, was put forward by Grégoire et al. (2015). This paper came up to the conclusion that the Logan and Tzelgov's conception was inappropriate for the MSE. However, we noted that there is a difference in conditions where note naming and word reading are practiced in real-world settings, which could prevent a direct generalization to reading. We wrote: "Musicians are exposed to printed music mainly, if not exclusively, on occasions where pronouncing the names of the notes vocally (e.g., at the beginning of practice) or subvocally (e.g., when playing an instrument later in practice) is well suited. By contrast, printed language is so ubiquitous in the environment of educated people that reading anything around us would be counterproductive. Optimal adaptation requires the skill of not attending to, or disengaging attention from, printed language when necessary. As a consequence, it remains possible that a part of the decrease in interference in Stroop tasks involving reading reflects the learned ability to ignore printed matters when the context makes reading inappropriate in real-world conditions. In this framework, improved cognitive control could indeed emerge as a function of practice, but as a domain-specific consequence of learning not to respond to stimuli whenever responding would distract from other activities" (pp. 424–425).

There is an additional reason that could explain that practice triggers a better control for reading and not for note naming in the present experiment, namely the fact that reading learning was more advanced than note-naming learning at the time of testing. A plausible hypothesis, indeed, is that a better control does not occur from the very beginning of training, but later on, let alone because engaging control processes would be objectless until some interference has begun to emerge.

At this stage, we may conclude that combining *Interference* and *Practice-related control* provides a satisfactory account of the data whenever *Interference* is applied for note naming and *Practice-related control* to word reading. This holds for the upwards or downwards directions of the effects. However, when the predicted amplitude of the effects is considered, the quality of fit appears far looser. Indeed, the drastic improvement in note-naming abilities observed between the

first and the fifth year of musical education is assumed to mediate the very tiny increase of the MSE over the same period, and conversely, the moderate improvement in word-reading abilities is assumed to generate the drastic decrease of the RMSE. Although there is no compelling reason to postulate that the magnitude of an effect must be proportional to the magnitude of the cause, the mismatch is here so striking that it strongly penalizes the potential action of *Practice-related control*.

4.3. Adding Resistance to interference

Rather than assuming that interference may decrease due to a change in the interfering process as in the prior interpretations, this last account posits that interference is moderated by practice-dependent changes of the *interfered* process, which would be increasingly immune to the interference generated by another process.

In this framework, the impressive decrease of the RMSE across the five years of musical training would be due to the strong gain in strength of note naming during the same period, and the consecutive rise of its resistance to interference. It is worthy to note, in addition, that the inhibitory power of word reading is quite consequent, which may explain the substantial size of the RMSE, but increases only moderately during the five years. The end-result is that the negative component of the effect (due to the immunity to interference of the to-be-interfered process) increases more abruptly than the positive component (due to the interfering power of the interfering process). The evolution of the MSE would be explained in opposite terms. The MSE would be stable or in slight augmentation because the resistance to interference of word reading, although strong from the outset (which may account for the small size of the MSE) increases only weakly during the period at hand. In addition, the inhibitory power of note naming begins at a low level, but increases substantially during the five years. Thus, in principle, taking into account the strength of the interfered process in addition to the strength of the interfering process could provide an interpretation for the whole pattern of results.

4.4. Conclusions

The observation of how Stroop interference evolves when children are trained in parallel to read and to name musical notes revealed an unexpected interactive pattern: the MSE, in which note naming is the interfering process, slightly increased, whereas the RMSE, in which word reading is the interfering process, exhibited a sharp decrease. A first major conclusion is that taken in isolation, the conventional and more common conception, according to which interference would be a direct function of the strength of the interfering process (coined here as *Interference*), is unable to account for this pattern. Indeed, the increase of the MSE appears to be overly smooth to be driven by the drastic improvement of note-naming abilities, and still more compellingly, the decrease of the RMSE is inconsistent with the observed improvement of word-reading abilities.

We listed three other processes opening to predictions about how the strength of interference evolves with training. Again, none of these processes was able to explain the interactive pattern when taken in isolation. We then examined how well the results are explained when *Interference* is complemented with, in succession, *Age-related control*, *Practice-related control*, and *Resistance to interference*. Considering jointly *Interference* and *Age-related control*, (e.g., Comalli et al., 1962; La Heij & Boelens, 2011) does not achieve to explain both the MSE and the RMSE in a satisfactory way. Indeed, the strong decrease of the RMSE would require a very strong effect of age-related control, which is not compatible with the data concerning the MSE. *Practice-related control* (Logan, 1982; Tzelgov et al., 1990) also generates a growing cognitive control, but thought of as linked to extensive practice in a specific task. Given that there are independent reasons for assuming that control increases only for word reading, this process was used as predictor for the RMSE, in which reading was the interfering process, whereas

interference was used as predictor for the MSE. This approach works reasonably well for predicting whether interference increases or decreases, but a large effect (the decrease of the RMSE) is attributed to the small improvement in word reading strength, and a small effect (the increase of the MSE) is attributed to a large improvement in note naming strength. Finally, *Resistance to interference* (e.g., Cohen et al., 1990), which makes performance increasingly resistant to the interference generated by another task, provides the best complement to *Interference* to account for the interactive pattern, because both the direction and the size of the effects are correctly predicted: large effects are mediated by large improvement in strength, while small effects are mediated by small improvement in strength.

To further examine this issue, simulations were run to assess whether introducing *Age-related control*, *Practice-related control*, and *Resistance to interference* in addition to *Interference*, significantly improves the predictions made from *Interference* alone on the size of the MSE on the one hand, and the RMSE on the other hand. We used a simple correlational approach, and in each case a test was performed to examine whether the resulting increase (if any) of the correlations was reliable. The details are reported in Appendix (and an Excel file comprising the step-by-step computations can be downloaded at <https://osf.io/2zdcw/>). In agreement with the qualitative analysis above, only the introduction of *Resistance to interference* significantly improved the predictions issued from *Interference* for both the MSE and the RMSE.

It is worth stressing that our study does not rule out the idea that cognitive control could improve either with age or with extensive practice in the task. We just suggest that these processes are neither necessary nor sufficient to account for the reported data. We do assert, however, that *Resistance to interference*, the growing immunity to interference of automatic responses, seems to be required. When considered in addition to the widespread idea that the amount of induced interference increases with the strength of the interfering process, the notion of resistance to interference of the interfered process leads to a full and parsimonious account of the interactive pattern of result. Considering Stroop interference as depending on these two sources of influence is a foundational principle of the influential Cohen et al. (1990) model, but this component of the model did not arouse much attention since then. The reason is certainly that in most Stroop-like procedures involving real-world abilities, the strength of the interfered process (e.g., color naming, object naming, numerical or spatial location knowledge) is very difficult to control and manipulate, hence making its role in the dynamics of the Stroop effect hard to detect.

The immunity to interference has not been ignored in the literature on automatism. However, this property has been investigated through other tasks, relying mainly on the dual-task paradigm. Because using different tasks to assess different properties does not allow the exploration of the interrelationships between these properties, the irrepressible nature of automatic behavior (which generates interference in concurrent tasks) and the propensity of automatic behavior to be immune with regard to the potential interference from these tasks, have never been considered jointly in empirical studies, and as a way of consequence, in theoretical appraisal. The musical Stroop paradigm provides a privileged opportunity to study the relationships between these properties.

Appendix A

Simulations were run to examine whether introducing *Age-related control*, *Practice-related control*, and *Resistance to interference* processes in addition to *Interference*, improves the predictions made from *Interference* alone on the size of the MSE on the one hand, and the RMSE on the other hand. We used a simple correlational approach, and in each case a test was performed to examine whether the resulting increase in the correlations was significant (the tests were performed with the paired.r function of the R package ‘psych’ assessing the difference between two nonindependent correlations).

The amount of interference generated by *Interference* was posited to be equal to the strength of the interfering process (recall that correlations are insensitive to any linear transformation, hence making most scale changes inconsequential). The strength of each pathway was assessed for each participant as the opposite of the mean RT collected in the relevant ability test. A proportion of the resulting value was removed, depending on different factors for *Age-related control* and *Resistance to interference*. An Excel file comprising the raw data and the details of the computations (with a worksheet for each account) can be downloaded at <https://osf.io/2zdcw/>

A.1. Adding Age-related control

The proportion of the interference generated by the competing process was assumed to depend on participants' age. The proportion to be subtracted was set to zero for the youngest child and to one for the oldest child. This procedure amounts to assume that the youngest child is unable to refrain interference, whereas, at the opposite the oldest child would be able to resist to any form of interference. This hypothesis is obviously an oversimplification, which maximizes the effect of age.⁴ The proportion to be subtracted for the other children was assessed as a linear function of age, as expressed in days. The final results were correlated with the observed values across the 100 participants. More precisely, the theoretical amount of interference as assessed from the strength of word reading, minored by cognitive control, was correlated with the observed RMSE, and likewise, the theoretical amount of interference as assessed from the strength of note naming, also minored by cognitive control, was correlated with the observed MSE.

The question was whether subtracting a part of the interference to simulate age-dependent cognitive control leads to better predictions than when only the strength of the interfering process was considered. For the RMSE, the response is mixed. While the initial correlation was negative ($r = -0.402$), the correlation after correction was positive ($r = 0.157$) and significantly different from the initial correlation, $t(97) = 4.99$, $p < .001$. However, the corrected correlation was not significantly different from zero ($p = .120$), which means that no reliable prediction is possible. For the MSE, the initial correlation was positive and significant ($r = 0.228$, $p = .022$). Subtracting a part for cognitive control has no substantial effect. The correlation after correction was numerically larger than before ($r = 0.265$), but the difference was not significant, $t(97) = 0.99$, $p = .326$. This suggests that taking into account the growing capacity of cognitive control in children has only limited consequences.

A.2. Adding Resistance to interference

Correlational analyses were run as for *Age-related control*, except that the values withdrawn from the interference generated by the interfering process, as inferred from the strength of this process, was no longer proportional to age, but instead proportional to the strength of the to-be-interfered process. For each child, a theoretical estimate of the RMSE was computed as the interference depending on the automaticity of word reading (the interfering process in the note-naming task), reduced by a proportion depending on the automaticity of note naming (the to-be-interfered process). The proportion of interference to be withdrawn was set to zero for the child with the smallest strength of note naming, and to one for the child with the highest strength (see Footnote 2 above), with the intermediate values being assessed as a

⁴ There is a range of other possibilities. For instance, the proportions may be set to 0.30 for the youngest child and 0.50 for the oldest child. All the possible combinations were explored (with a 0.10 resolution). Only the correlations observed with the maximal range of variation (i.e., 0 to 1) are reported here and below, because other values result in effects going in the same direction, although unsurprisingly smaller.

linear function of strength. Over the participants, the correlation of this score with the observed RMSE was $r = 0.243$. This value remains moderate in size, but is significant ($p = .014$), and differs from the correlation observed when only the strength of the interfering process is considered, $t(97) = 7.80$, $p < .001$. A theoretical estimate of interference was computed for the MSE in the same way, except that the roles of word reading and note naming were reversed. Over the participants, the correlation of this score with the observed MSE was $r = 0.311$, $p = .002$. This value is reliably greater than the correlation observed when only the strength of the interfering process is considered, $t(97) = 2.79$, $p = .006$. To sum up, considering the strength of the to-be-interfered process in addition to the strength of the interfering process improves correlations for both the RMSE and the MSE, and allows reliable predictions for the two effects.

A.3. Adding Practice-related control

In the account proposed in the main text, the *Practice related control* would not work for the MSE, because there is no need for musicians to improve their control over note naming. As a consequence, there is no change for the correlation with regard to *Interference* alone, i.e., $r = 0.228$. By contrast, the *Practice related control* would apply to the RMSE, because anyone needs to improve control over word reading. As a consequence, the RMSE would be inversely proportional to the strength of word reading. The end-result is that the sign of the initial correlation is now inverted (namely, the initially negative correlation becomes positive, $r = 0.402$).

References

- Akiva-Kabiri, L., & Henik, A. (2012). A unique asymmetrical Stroop effect in absolute pitch possessors. *Experimental Psychology*, 59(5), 272–278. <https://doi.org/10.1027/1618-3169/a000153>.
- Arend, I., & Henik, A. (2015). Choosing the larger versus choosing the smaller: Asymmetries in the size congruity effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(16), 1821–1830. <https://doi.org/10.1037/xlm0000135>.
- Armengol, C. G. (2002). Stroop test in Spanish: Children's norms. *The Clinical Neuropsychologist*, 16(1), 67–80. <https://doi.org/10.1076/clin.16.1.67.8337>.
- Baker, M. (2016). Is there a reproducibility crisis? *Nature*, 533(7604), 452–454. <https://doi.org/10.1038/533452a>.
- Bedard, A. C., Nichols, S., Barbosa, J. A., Schachar, R., Logan, G. D., & Tannock, R. (2002). The development of selective inhibitory control across the life span. *Developmental Neuropsychology*, 21(1), 93–111. https://doi.org/10.1207/S15326942dn2101_5.
- Ben-Shalom, T., Berger, A., & Henik, A. (2013). My brain knows numbers! An ERP study of preschoolers' numerical knowledge. *Frontiers in Psychology*, 4, 716. <https://doi.org/10.3389/fpsyg.2013.00716>.
- Blais, C., & Besner, D. (2006). Reverse Stroop effects with untranslated responses. *Journal of Experimental Psychology: Human Perception and Performance*, 32(6), 1345–1353. <https://doi.org/10.1037/0096-1523.32.6.1345>.
- Blais, C., & Besner, D. (2007). A reverse Stroop effect without translation or reading difficulty. *Psychonomic Bulletin & Review*, 14(3), 466–469. <https://doi.org/10.3758/BF03194090>.
- 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(3), 332–361. <https://doi.org/10.1037/0033-295X.97.3.332>.
- Comalli, P. E., Jr., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood, and aging. *Journal of Genetic Psychology*, 100, 47–53. <https://doi.org/10.1080/00221325.1962.10533572>.
- Dash, J., & Dash, S. (1982). Cognitive developmental studies of the Stroop phenomena: Cross-sectional and longitudinal data. *Indian Psychologist*, 1(1), 24–33.
- Dunbar, K., & MacLeod, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance*, 10(5), 622–639. <https://doi.org/10.1037/0096-1523.10.5.622>.
- Durgin, F. H. (2003). Translation and competition among internal representations in a reverse Stroop effect. *Perception & Psychophysics*, 65(3), 367–378. <https://doi.org/10.3758/BF03194568>.
- Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance*, 8(6), 875–894. <https://doi.org/10.1037/0096-1523.8.6.875>.
- Gliksman, Y., Itamar, S., Leibovich-Raveh, T., Melman, Y., & Henik, A. (2016). Automaticity of conceptual magnitude. *Scientific Reports*, 6, 21446. <https://doi.org/10.1038/srep21446>.
- Goldfarb, L., & Henik, A. (2007). Evidence for task conflict in the Stroop effect. *Journal of Experimental Psychology: Human Perception and Performance*, 33(5), 1170–1176. <https://doi.org/10.1037/0096-1523.33.5.1170>.
- Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The musical Stroop effect. Opening a new avenue to research on automatism. *Experimental Psychology*, 60(4), 269–278. <https://doi.org/10.1027/1618-3169/a000197>.
- Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2014). About the unidirectionality of interference: Insight from the musical Stroop effect. *The Quarterly Journal of Experimental Psychology*, 67(11), 2071–2089. <https://doi.org/10.1080/17470218.2014.896932>.
- Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2015). How does Stroop interference change with practice? A reappraisal from the musical Stroop paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(2), 417–425. <https://doi.org/10.1037/xlm0000044>.
- Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition*, 10(4), 389–395. <https://doi.org/10.3758/BF03202431>.
- Klenberg, L., Korkman, M., & Lahti-Nuutila, P. (2001). Differential development of attention and executive functions in 3- to 12-year-old Finnish children. *Developmental Neuropsychology*, 20(1), 407–428. https://doi.org/10.1207/S15326942dn2001_6.
- La Heij, W., & Boelens, H. (2011). Color-object interference: Further tests of an executive control account. *Journal of Experimental Child Psychology*, 108(1), 156–169. <https://doi.org/10.1016/j.jecp.2010.08.007>.
- Logan, G. D. (1982). On the ability to inhibit complex movements: A stop-signal study of typewriting. *Journal of Experimental Psychology: Human Perception and Performance*, 8(6), 778–792. <https://doi.org/10.1037/0096-1523.8.6.778>.
- Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology*, 39(2), 367–386. <https://doi.org/10.1037/h0080066>.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2), 163–203. <https://doi.org/10.1037/0033-2909.109.2.163>.
- 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(1), 126–135. <https://doi.org/10.1037/0278-7393.14.1.126>.
- Martin, M. (1978). Speech recoding in silent reading. *Memory & Cognition*, 6(2), 108–114. <https://doi.org/10.3758/BF03197435>.
- Melara, R. D., & Mounts, J. R. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition*, 21(5), 627–645. <https://doi.org/10.3758/BF03197195>.
- Morton, J., & Chambers, S. M. (1973). Selective attention to words and colours. *The Quarterly Journal of Experimental Psychology*, 25(3), 387–397. <https://doi.org/10.1080/14640747308400360>.
- Palef, S. R., & Olson, D. R. (1975). Spatial and verbal rivalry in a Stroop-like task. *Canadian Journal of Psychology*, 29(3), 201–209. <https://doi.org/10.1037/h0082026>.
- Peru, A., Faccioli, C., & Tassinari, G. (2006). Stroop effects from 3 to 10 years: The critical role of reading acquisition. *Archives Italiennes de Biologie*, 144(1), 45–62.
- Rand, G., Wapner, S., Werner, H., & McFarland, J. H. (1963). Age differences in performance on the Stroop color-word test. *Journal of Personality*, 31(4), 534–558. <https://doi.org/10.1111/j.1467-6494.1963.tb01318.x>.
- Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture-word interference task. *Child Development*, 46(1), 247–253. <https://doi.org/10.2307/1128859>.
- Schadler, M., & Thissen, D. M. (1981). The development of automatic word recognition and reading skill. *Memory & Cognition*, 9(2), 132–141. <https://doi.org/10.3758/BF03202327>.
- Schiller, P. H. (1966). Developmental study of color-word interference. *Journal of Experimental Psychology*, 72(1), 105–108. <https://doi.org/10.1037/h0023358>.
- Stanovich, K. E., Cunningham, A. E., & West, R. F. (1981). A longitudinal study of the development of automatic recognition skills in first graders. *Journal of Reading Behavior*, XIII(1), 58–59. <https://doi.org/10.1080/10862968109547394>.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662. <https://doi.org/10.1037/h0054651>.
- Tzelgov, J., Henik, A., & Leiser, D. (1990). Controlling Stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(5), 760–771. <https://doi.org/10.1037/0278-7393.16.5.760>.