THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2014
http://dx.doi.org/10.1080/17470218.2014.896932

About the unidirectionality of interference: Insight from the musical Stroop effect

Laurent Grégoire, Pierre Perruchet, and Bénédicte Poulin-Charronnat
LEAD CNRS-UMR 5022, University of Bourgogne, Dijon, France

The asymmetry of interference in a Stroop task usually refers to the well-documented result that incongruent colour words slow colour naming (Stroop effect) but incongruent colours do not slow colour word reading (no reverse Stroop effect). A few other studies have suggested that, more generally, a reverse Stroop effect can be occasionally observed but at the expense of the Stroop effect itself, as if interference was inherently unidirectional, from the stronger to the weaker of the two competing processes. We describe here a situation conducive to a pervasive mutual interference effect. Musicians were exposed to congruent and incongruent note name/note position patterns, and they were asked either to read the word while ignoring the location of the note within the staff, or to name the note while ignoring the note name written inside the note picture. Most of the participants exhibited interference in the two tasks. Overall, this result pattern runs against the still prevalent model of the Stroop phenomenon [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]. However, further analyses lend support to one of the key tenets of the model, namely that the pattern of interference depends on the relative strength of the two competing pathways. The reasons for the impressive differences between the results collected in the present study and in the standard colour–word (or picture–word) paradigms are also examined. We suggest that these differences reveal the importance of stimulus–response contingency in the formation of automatisms.

Keywords: Stroop; Interference; Automatism; Musical expertise; Contingency.

Grégoire, Perruchet, and Poulin-Charronnat (2013) devised a new version of a musical Stroop task (for an earlier musical Stroop paradigm investigating motor automatisms in pianists, see Stewart, Walsh, & Frith, 2004). The basic arrangement comprises a staff with a note in various positions (see Figure 1). A name of a note is printed inside the note. For the congruent condition (Figure 1a), the note name is congruent with the note position on the staff, whereas in the incongruent condition (Figure 1b), note name and position are incongruent. Musicians asked to read the written names of the notes showed impaired processing in the incongruent condition compared to the congruent condition. This *musical Stroop effect* (MSE) attests to the interference[1] generated by the automaticity of note naming in musicians.

The primary motivation for studying the automaticity of note naming, instead of the automaticity of word reading as do the conventional

Correspondence should be addressed to Laurent Grégoire, LEAD-CNRS UMR5022, Université de Bourgogne, Pôle AAFE, 11 Esplanade Erasme, 2100 Dijon, France. E-mail: laurent.gregoire@u-bourgogne.fr

This research was supported by the Agence Nationale de la Recherche [grant number ANR-09-JCJC-0129]; the Conseil Régional de Bourgogne [grant number 2012-9201AAO05S01353].

[1]For the sake of simplicity, except when specified otherwise, the difference between incongruent and congruent conditions is coined here as interference, even though a part of the effect may arguably be due to facilitation in the congruent condition.

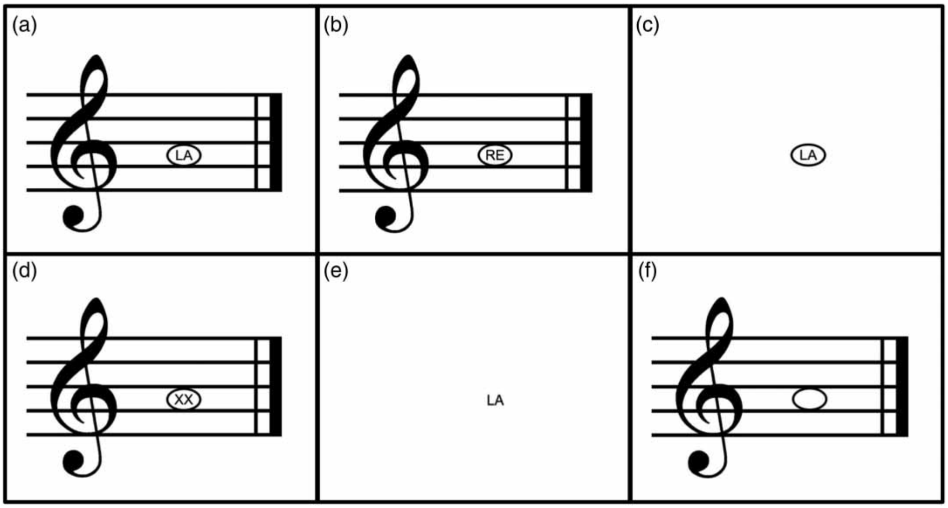

Figure 1. *Examples of the different conditions used in the experiment: (a) congruent condition; (b) incongruent condition; (c) neutral condition for word reading; (d) neutral condition for note naming; (e) reading-ability test; (f) note-naming ability test. Note that in the musical French notation (and 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.*

version of the Stroop paradigm and most Stroop-like tasks (e.g., Flowers, Warner, & Polansky, 1979; Glaser & Düngelhoff, 1984; Virzi & Egeth, 1985), stemmed from the greater flexibility in manipulating practice. Of particular interest is the fact that practice level can be decoupled from age and reading-skill abilities, hence offering to researchers the possibility of manipulating a host of new variables (Grégoire, Perruchet, & Poulin-Charronnat, 2014). However, other questions are linked to the fact that, in the MSE, reading is not put outside overall; its status is just reversed, from the status of the interfering process to the status of the interfered process. A Stroop effect in which reading is interfered by another process is often coined as a *reverse Stroop effect* (RSE).[2] In the remainder of this introduction, we first outline the main findings regarding the RSE, as well as their theoretical implications, and then we examine why the Grégoire et al.'s (2013) musical Stroop is a unique test-bed to address one of the major issues raised by this literature.

The reverse Stroop effect

Stroop (1935) was the first to examine whether *reading* colour words could be impaired by an incongruent colour–word combination. He reported an RSE, but only after extensive practice to name the colour of incongruent words. In addition, this effect was short-lived, as evidenced by the disappearance of the RSE a few days after the participants stopped practising. Dulaney and Rogers (1994) replicated this result and showed that the RSE obtained after practising the colour-naming task for incongruent words was due to the development of a "reading suppression response". Participants seemingly did not automatize colour naming but instead acquired a transitory capacity to inhibit the reading response. This conclusion was supported by Ellis, Woodley-Zanthos, Dulaney, and Palmer (1989), who observed that participants previously trained to name the colour of control items (a series of coloured Xs), a task that offers no opportunity to inhibit reading, gave

[2]This terminology is conventional (e.g., MacLeod, 1991), but not without its own shortcomings, notably because it applies only to situations in which reading is one of the two competing processes. Given that an overwhelming proportion of the Stroop literature involves reading, however, this is not a strong limit, and this definition is used throughout this paper.

no evidence for an RSE. More recently, MacLeod (1998) failed to observe an RSE with the now standard single-item procedure despite very extensive training to name incongruent colour words. He attributed the fleeting effect reported in earlier studies to the use of the multiple-item version of the task, in which trial types are not mixed. To sum up, the pervasive failure to get an RSE under usual reading conditions has led to construe the asymmetry of interference as a ubiquitous property of the reading processes.

The overall picture, however, is more complex. In fact, an RSE has been obtained in colour–word Stroop versions when the usual conditions of reading were strongly degraded (e.g., Dunbar & MacLeod, 1984; Melara & Mounts, 1993), or still when verbal responses were replaced by motor responses (e.g., Blais & Besner, 2006, 2007; Durgin, 2003; Melara & Mounts, 1993). Moreover, other studies (e.g., Akiva-Kabiri & Henik, 2012; Palef & Olson, 1975) also observed an RSE in Stroop-like procedures that involved a reading task but another competing process other than colour naming. At first glance, these data could simply suggest that Stroop asymmetry is not as ubiquitous as once thought. But a closer scrutiny of the few studies investigating concurrently the RSE and the standard effect completes this conclusion with a more intriguing observation.

Let us consider the recent study of Akiva-Kabiri and Henik (2012). In this study, musician participants were asked to read the name of notes while hearing a tone that could either correspond to the note (congruent condition) or not (incongruent condition). Absolute pitch possessors (a mainly inherited ability) showed a significant RSE. Reading was impaired when the written word was incongruent with the tone. By contrast, those participants did not show interference when they were asked to label the auditory tone in the presence of an incongruent written note name. Moreover, control musicians without absolute pitch showed the inverse pattern: They were not affected by an incongruent tone when reading a note name, but they were affected by an incongruent written note name when labelling the tone. In a nutshell, the results from Akiva-Kabiri and Henik suggest that, by and large, the Stroop effect and the RSE are mutually exclusive.

A few earlier studies suggested similar conclusions. Palef and Olson (1975) used a paradigm in which the words *above* and *below* were presented either above or below the fixation point. In their first experiment, reading the word showed interference from incongruent locations, hence evidencing an RSE. Strikingly, however, incongruent word meanings did not interfere with decisions about spatial position. The authors reasoned that this asymmetrical effect could be due to the fact that spatial position was processed faster than word meaning. To test this hypothesis, they designed a second experiment in which the relative duration required for each task was reversed. The words were now displayed at the fixation point, with an asterisk serving as reference to the spatial judgement being either above or below the words. Under these conditions, decisions about spatial position took longer than reading the words, and, concurrently, the direction of interference was reversed. In other words, everything happened as if interference could act in either direction, but with a single dimension at a time as a function of conditions. Some amount of interference was bidirectional, however, but only in a subgroup of participants in Experiment 2 for whom, due to the chronological ordering of the tasks, the processing times turned out to be approximately equal between reading and judgements of spatial position.

A similar pattern of data emerged from Melara and Mounts's (1993) experiments with the standard colour–word Stroop paradigm. When the readability of the colour word was reduced (small fonts were read a long way away from the screen), the authors observed a large RSE on both response times and errors, but at the expense of the standard effect, which did not reach significance (Experiment 2). Thus the usual pattern of asymmetry was inverted. In two other experiments (Experiments 1 and 4), the authors equalized the perceptual discriminability of words and colours, again by shrinking the visual angle of the stimuli (but to a lesser extent than in Experiment 2). Under these very specific conditions, both the standard effect and its reverse were observed, but these effects were small and volatile. To borrow the authors' words: "subjects displayed meager, fleeting, and generally symmetrical effect of irrelevant variation" (Melara & Mounts, 1993, p. 642).

The final picture is that the asymmetry of the Stroop effect cannot be understood as the ubiquitous prevalence of reading over any other process, even if it is its most usual expression. However, the asymmetry could be redefined as the fact that whatever the two competing processes, interference may go in either direction as a function of participants and experimental conditions, but, crucially, in a single direction at a time.[3] A few studies reported statistically significant bidirectional interference, but with effects that were small, ephemeral, and circumscribed to very specific experimental arrangements. As pointed out by Cohen, Dunbar, and McClelland (1990), this pattern of data, based on group statistics, may reflect a mixture of unidirectional effects going in the opposite direction for different subjects and therefore cannot be taken as a decisive argument to refute the principle of strict unidirectionality.

Theoretical implications

Cohen et al.'s (1990) connectionist model is still acknowledged as one of the prevalent models of the Stroop effect (e.g., Blais & Besner, 2006, 2007; MacLeod & MacDonald, 2000; Protopapas, Archonti, & Skaloumbakas, 2007; Roelofs, 2005). Cohen et al. did not refer to Palef and Olson (1975), and they were obviously unaware of later studies (Akiva-Kabiri & Henik, 2012; Melara & Mounts, 1993). However, one of the main motivations for the development of their model was the report of related data by MacLeod and Dunbar (1988). Although MacLeod and Dunbar's paradigm did not involve reading, and hence does not concern the RSE as defined above, their results are highly relevant for our concern. In MacLeod and Dunbar, a specific colour name was arbitrarily assigned to each of four black-and-white polygons, and participants were instructed to learn these names along successive practice sessions. After different amounts of training, shapes were presented in colours that were either congruent or incongruent with their name. Participants carried out two tasks: They had to name the shapes ignoring their colours and to name the colours disregarding the shapes. Early in training, colour had an effect on shape naming but not vice versa. After extensive training in shape naming, the opposite pattern of results was observed: Shape had an effect on colour naming but not vice versa. Therefore, the effect and its reverse were mutually exclusive, at least when the extreme ends on the level of practice were considered (i.e., Day 1 and Day 20 in their Experiment 3).

How does Cohen et al.'s (1990) model account for MacLeod and Dunbar's (1988) results, and notably for the fact that the direction of interference is reversed even though the strength of the colour-naming pathway remains, presumably, constant? In this model, the attributes of automaticity depend on the strength of a processing pathway, which is itself a function of the amount of training received in the task at hand. The critical point is that when two pathways conflict, the resulting pattern of interference does not depend on the absolute strength of each pathway, but on the *relative* strength of the two pathways. To quote Cohen et al., it is "the relative strength, compared with a competing pathway [that is] important in determining whether a process will produce or be subject to interference in a Stroop-like task" (p. 348). Figure 2 illustrates how the direction of interference between two processes, A and B, may be reversed even though the strength of B does not vary (Situations 1 vs. 4 in the figure).

Nevertheless, there is some discrepancy between MacLeod and Dunbar's (1988) empirical data and the model's predictions, regarding what occurs

[3]Using two logographic scripts, Japanese and Chinese, Verdonschot, La Heij, and Schiller (2010) reported standard Stroop effects in a picture–word interference task. However, when investigating the RSE, they observed that reading Japanese kanji was *shortened* by the incongruent pictures compared to unrelated pictures. No effect at all was observed with Chinese hànzì. Irrespective of the interpretation of these findings, they concur to generalize the asymmetry of the Stroop effect to nonalphabetic writing systems.

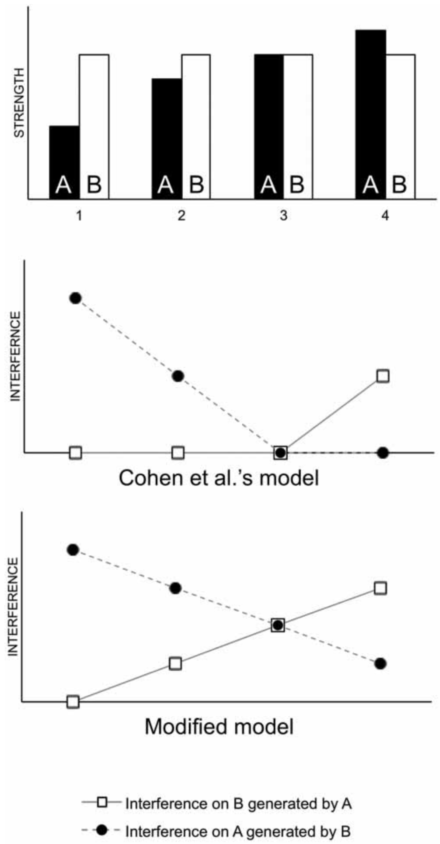

Figure 2. *Schematic predictions of Cohen et al.'s (1990) model and its modified version, in four situations, in which the strength of Process B is kept constant, while the strength of Process A gradually increases. In both models, the amount of interference (on the y-axis, arbitrary units) depends on the relative strength of the competing processes. However, in Cohen et al.'s model, only the stronger of the two processes generates interference, giving unidirectional interference when the strength of the two processes differs and no interference at all when the two processes have equal strength. The modified, "softened" model allows some amount of interference from the weaker process on the stronger process.*

when the two competing processes are of equal strength. In MacLeod and Dunbar, each dimension influenced the other to the same extent after an intermediate amount of practice (i.e., Day 5). In the model, there is no possibility for a process A to interfere with a process B (this would imply A > B) and to be interfered with by a process B (this would imply A < B), and reciprocally. As a consequence, the amount of mutual interference observed in the simulations when the two processes were of comparable strength was negligible (see Figure 2, Situation 3). Cohen et al. (1990) themselves acknowledge that their difficulty to account for the mutual interference reported by MacLeod and Dunbar is a robust property of their model, and they construed this outcome as a shortcoming (p. 353).

Is this shortcoming really devastating for the validity of Cohen et al.'s (1990) model? One may reasonably think that a single empirical demonstration of bidirectional interference (MacLeod & Dunbar, 1988) does not strike a fatal blow to the model. Unfortunately, as shown in the prior section, the other relevant studies do not clearly confirm or disprove this demonstration. A small amount of mutual interference is sometimes reported in quite specific conditions, a form of evidence that moderately challenges Cohen et al.'s model without providing a definitive case against it. The present study was devised to clear up the current ambiguities.

THE PRESENT STUDY

As described above, replacing colours with musical notes as the irrelevant dimension in a reading task resulted in the observation of a reliable RSE in musicians (Grégoire et al., 2013). Reading times were slowed down by incongruent note positions, even though note names were not degraded and were printed in a font that certainly surpassed in size the fonts used in most books or newspapers. However, as such, these results provide only weak arguments against the unidirectionality of interference. First, the effects were small in amplitude (within a 8–10-ms range).[4] The first aim of the present study was to replicate and strengthen Grégoire et al.'s (2013) earlier evidence for an

[4]In addition, Zakay and Glickson (1985) used similar conditions as a part of a more general paradigm, and verbal responses showed no reliable evidence for an MSE. However, methodological limitations are probably responsible for this failure, as analysed in Grégoire et al. (2013).

MSE, by using a somewhat simplified procedure that is described later. To anticipate, the results clearly fulfilled our expectations in this regard. Second and above all, Grégoire et al. provided no evidence for mutual interference, given that their experiments did not include a task in which reading would have been the irrelevant dimension.

In the following experiments, musicians were exposed to the congruent and incongruent note name/note position patterns used in Grégoire et al. (2013), but they received two types of instructions in succession: In addition to being asked to read the word while ignoring the location of the note within the staff, as in Grégoire et al., musicians were asked to name the note while ignoring the note name written inside the note picture. This latter condition was devised to generate a "reverse MSE" (hereafter: RMSE).[5] In addition, to assess the relative strength of the processing pathways, musicians were submitted to a reading-ability test and a note-naming ability test after the experimental sessions. The note-naming test was also exploited to check whether all musicians reached a sufficient level of musical expertise.

Method

Participants

All participants were undergraduate psychology students at the University of Bourgogne, who received course credit for their participation. They were French native speakers and reported normal or corrected vision. Twenty-six participants had received formal musical training and had played a musical instrument for at least five years. Four of them were removed from the analyses on the basis of their performance in the note-naming ability test (see detail below). The remaining 22 participants (17 women, 5 men) had an average of 12.50 ($SD = 4.51$) years of musical training. Twenty-two other participants had never studied or practised music (18 women, 4 men).

Materials

The experimental material was composed of four types of stimuli. For the congruent and incongruent conditions, stimuli consisted of a treble staff with a note picture, which could appear on each of the 13 possible positions going from C4 to A5. 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 (Figure 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 could be LA, SI, RE, MI, FA, or SOL; Figure 1b).

The neutral condition was aimed at assessing whether the overall difference between congruent and incongruent trials was due to interference or facilitation. For the word-reading task, a note name was inserted inside a note picture, and the resulting pattern was displayed on the same 13 spatial locations as those for congruent and incongruent trials (from C4 to A5). However, the staff was not represented (Figure 1c). The stimuli described so far were identical to those used in Grégoire et al. (2013). For the note-naming task, a note picture was presented in a treble staff, with *XX* or *XXX* written inside (Figure 1d). Note pictures were also displayed at all 13 note locations.

Grégoire et al. (2013) also used stimuli that were made up of words that were not names of notes (and hence neither congruent nor incongruent with note locations), but instead highly frequent French words that were matched in length with note names. The results suggested that these additional conditions were not necessary in further investigations, and the authors conjectured that using only the experimental conditions involving note names would result in larger congruity effect. Indeed, as a by-product of mixing note names with other words, the proportion of congruent trials was reduced (20%), and it has been shown in the standard colour–word version that

[5] Given that the MSE is an instance of RSE as defined above, the RMSE, in which reading is the interfering process, has the status of a Stroop-like effect. This terminological quibble is unfortunate, but seems unavoidable, because it is difficult to revisit a terminology forged through eight decades of tradition during which reading has been the almost exclusive target in Stroop studies.

interference increases with the proportion of congruent trials (Lowe & Mitterer, 1982). In the present experiment, only note names were used.

The stimuli used in the independent tests of reading and note-naming abilities were designed to be closer than those involved in everyday situations. They are represented in Figure 1e and Figure 1f, respectively.

To prevent the iconic memory of the staff from influencing 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 centre 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 and Xs appeared in a font that roughly corresponds to the "Arial 14", uppercase font. The treble staff was 7.7 cm wide by 5.1 cm high.

Procedure

Musicians 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 participants. In the word-reading task, participants had to read aloud the printed word while ignoring the note picture. In the note-naming task, musicians were asked to name the note while ignoring the word written inside. Nonmusicians did not perform the note-naming task.

For each task, there were three mixed conditions: congruent (Figure 1a), incongruent (Figure 1b), and neutral (Figures 1c and 1d). For each condition, the stimuli appeared six times on each of the 13 locations, resulting in 78 trials per condition and 234 trials (78 × 3) for each task. On each trial, a fixation cross was displayed for 1 s at the centre of the screen before the appearance of the stimulus, which stayed on the screen until the participant's response. The interval between the response and the next trial was 1 s. The trials were pseudorandomly ordered for each participant, excluding immediate repetitions of the same response. They were displayed as six blocks of 39 trials each with a self-paced break between blocks.

For the musicians, 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 participants had to read note names (Figure 1e). The other test was a note-naming ability test, in which participants had to name notes (Figure 1f). Each test included 78 trials. The trials were pseudorandomly ordered for each participant, excluding immediate repetitions of words and notes. They were displayed as two blocks of 39 trials each with a self-paced break between blocks.

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 the error responses and the voice-key dysfunctions. After the experiment, the musicians filled out a questionnaire about their musical training.

Results

Ability tests (musicians)

Four musicians were removed from the analyses on the basis of their poor performance in the note-naming ability test: Three of them took longer than 1000 ms on average to name a note, and a last one made seven errors (a majority of the remaining 22 musicians made no error, and their highest error score was four).

Voice-key dysfunctions led to exclude 3.06% of the data. Unsurprisingly, there was virtually no error in the reading-ability test, $M = 0.12\%$, $SD = 0.39$. The proportion of errors in the note-naming ability test was significantly higher, $M = 1.47\%$, $SD = 2.05$, $t(21) = 3.10$, $p = .005$, $d = 0.94$. RTs for correct responses beyond three standard deviations of the overall mean per participant (0.73%) were removed. Correct RTs were also significantly shorter for word reading than for note naming, $M = 615.37$, $SD = 62.30$ versus $M = 742.28$, $SD = 94.33$, respectively, $t(21) = 8.21$, $p < .001$, $d = 1.79$. Note that there was no evidence for a speed–accuracy trade-off in the note-naming task, as shown by the positive (although nonsignificant) correlation between errors and RTs, $r(20) = .21$, $p = .35$.

Word reading (MSE)

Voice-key dysfunctions led to exclusion of 4.28% of the data. While nonmusicians made no reading errors, a small proportion of errors was observed for musicians (0.75%). An analysis of variance (ANOVA) on musicians' reading errors performed with condition (congruent, incongruent, neutral) as a within-subject variable showed a significant main effect of condition, $F(2, 42) = 4.88$, $p = .012$, $\eta_p^2 = .189$. The rate of errors was significantly higher in the incongruent condition ($M = 1.89\%$, $SD = 3.76$) than in the congruent condition ($M = 0.06\%$, $SD = 0.27$), $t(21) = 2.30$, $p = .032$, $d = 0.50$, attesting to an MSE. The rate of errors in the neutral condition ($M = 0.30\%$, $SD = 0.68$) was both lower than the rate of errors in the incongruent condition, $t(21) = 2.11$, $p = .047$, $d = 0.46$, and higher than the rate of errors in the congruent condition, $t(21) = 2.16$, $p = .042$, $d = 0.47$.

RTs for correct responses beyond three standard deviations of the overall mean per participant (0.59%) were removed. The remaining data are shown in Figure 3. An ANOVA on RTs was carried out with condition (congruent, incongruent, neutral) as a within-subject variable and musical expertise (musicians, nonmusicians) as a between-subject variable. There was no main effect of musical expertise, $F(1, 42) = 0.52$, $p = .474$, a significant effect of condition, $F(2, 84) = 52.94$, $p < .001$, $\eta_p^2 = .558$, and, more importantly, a significant Condition × Musical Expertise interaction, $F(2, 84) = 4.54$, $p = .013$, $\eta_p^2 = .098$.

For musicians, RTs were significantly longer in the incongruent condition than in the congruent condition, $t(21) = 3.21$, $p = .004$, $d = 0.70$. While the MSE reported by Grégoire et al. (2013) was in the 8–10-ms range, the difference reached 31.36 ms in this experiment, presumably due to the removing of control conditions. Two questions arise regarding the generality of the effect: First, is it observed for all participants? And second, for a given participant, is it observed across the entire latency distribution of his or her responses? Regarding the first issue, RTs were numerically longer in the incongruent condition than in the congruent condition for 20 out of the 22 musician participants. The second issue was addressed by

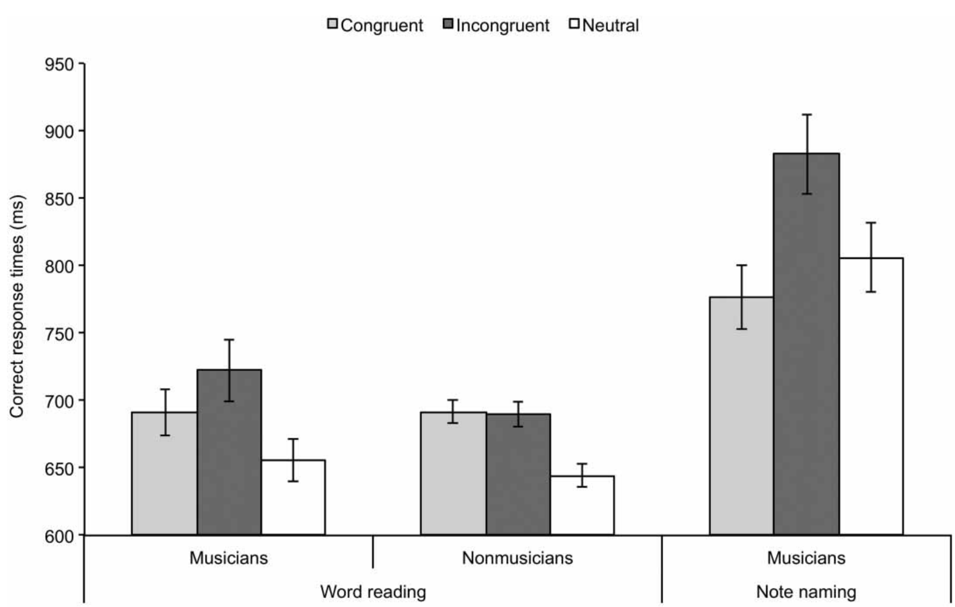

Figure 3. *Correct response times as a function of task, condition, and musical expertise. Error bars indicate standard errors.*

calculating so-called cumulative Vincentized distribution functions (Ratcliff, 1979; Roelofs, 2008). The rank-ordered RT distribution was divided into deciles (10% quantiles) for each participant separately for the congruent and the incongruent conditions, then the data were averaged across participants. The resulting curves are reported in Figure 4, upper panel. The MSE, which is

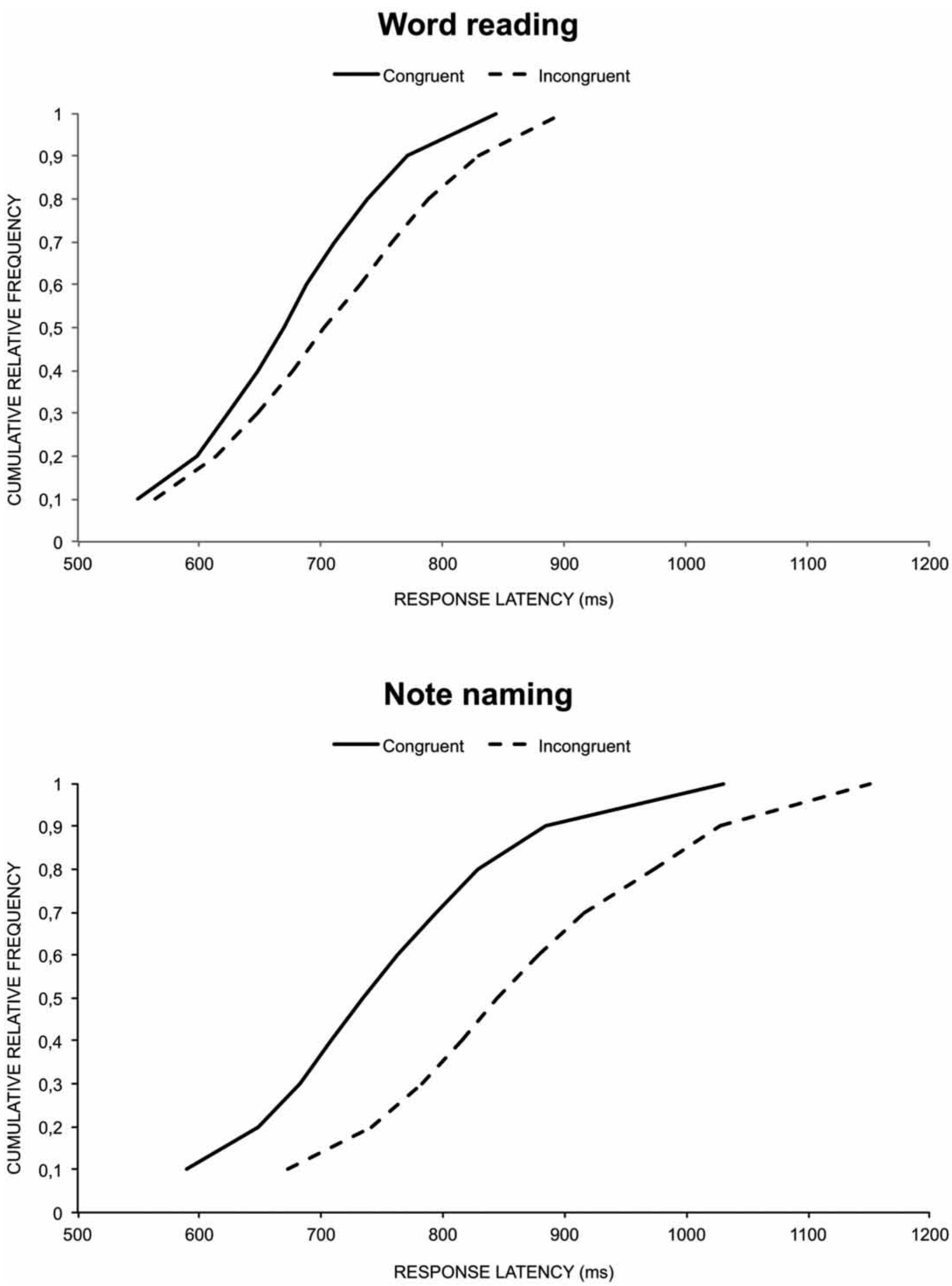

Figure 4. *Vincentized cumulative distribution curves for word reading and note naming in the congruent and incongruent conditions.*

represented on the figure as the horizontal distance between the curves, was stronger when RTs were longer, generating a significant interaction between congruity (congruent, incongruent) and deciles (1–10), $F(9, 189) = 10.95$, $p < .001$, $\eta_p^2 = .343$. Importantly, despite this effect, individual t-tests showed that the effect of conditions was significant even for the shortest RTs (first decile: $p = .045$, second decile: $p = .031$, other deciles: all $p < .01$).

RTs did not differ between incongruent and congruent conditions for the nonmusicians, $t(21) = 1.06$, $p = .301$. Comparing the effect between musicians and nonmusicians revealed a significant difference, $t(42) = 3.36$, $p = .002$, $d = 1.04$, attesting that the MSE was specific to musicians.

Grégoire et al. (2013) showed that the neutral word condition is not appropriate to provide a direct measure of facilitation and interference, because perceptual complexity is lower in this condition than in the congruent and incongruent conditions. Nevertheless, the effect of perceptual complexity can be assessed in nonmusicians, as a difference between the congruent and incongruent conditions (which gave nearly identical RTs) on the one hand, and the neutral condition on the other hand. The resulting value was 46.26 ms. If one supposes that musicians and nonmusicians are equally sensitive to perceptual complexity, this value can be added to the value observed for musicians in the neutral condition to obtain a neutral value that incorporates the effect of perceptual complexity. This corrected neutral value would thus be 701.95 ms (655.69 + 46.26). Given that the RTs in congruent and incongruent conditions for musicians were 691.16 ms and 722.52 ms, respectively, the overall difference (31.36 ms) may be decomposed into a facilitation effect (701.95 − 691.16 = 10.79 ms) and a larger interference effect (722.52 − 701.95 = 20.57 ms). Note that this analysis rests on the postulate that musicians and nonmusicians are equally sensitive to perceptual complexity, which is somewhat unrealistic. Most probably, musicians are less sensitive than nonmusicians to the complexity of a picture they are highly familiar with. This implies that the corrected neutral value should be lower, and hence that the part of facilitation in the analysis above is certainly overestimated. This descriptive analysis confirms Grégoire et al.'s observation that the MSE is mainly due to interference in the incongruent condition.

Note naming (RMSE)

Voice-key dysfunctions led to exclusion of 5.63% of the data. An ANOVA on naming errors performed with condition (congruent, incongruent, neutral) as a within-subject variable (recall that only musicians performed this task) showed a significant main effect of condition, $F(2, 42) = 32.13$, $p < .001$, $\eta_p^2 = .605$. Musicians made significantly more errors in the incongruent condition ($M = 7.10\%$, $SD = 4.56$) than in the congruent condition ($M = 0.86\%$, $SD = 1.79$), $t(21) = 6.03$, $p < .001$, $d = 1.32$, and than in the neutral condition ($M = 1.64\%$, $SD = 2.92$), $t(21) = 5.80$, $p < .001$, $d = 1.27$. Although there were more errors in the neutral condition than in the congruent condition, the difference failed to reach significance, $t(25) = 1.75$, $p = .095$, $d = 0.38$.

Correct RTs beyond three standard deviations of the overall mean per participant (1.21%) were removed. The remaining data are shown in Figure 3, right panel. As for the errors, an ANOVA carried out on the correct RTs with condition (congruent, incongruent, neutral) as a within-subject variable revealed a significant main effect of condition, $F(2, 42) = 64.49$, $p < .001$, $\eta_p^2 = .754$. RTs in the incongruent condition were significantly longer than those in the congruent condition, $t(21) = 9.20$, $p < .001$, $d = 2.01$, hence attesting to the presence of an RMSE. This effect was present for all musician participants, and cumulative Vincentized distribution functions calculated as above (Figure 4, lower panel) showed that the effect was observed throughout the entire RT range ($p < .001$ for all deciles). As for the MSE, the RMSE was stronger when RTs were longer, generating a significant Congruity × Decile interaction, $F(9, 189) = 9.38$, $p < .001$, $\eta_p^2 = .309$. In comparison with the neutral condition, we observed an interference in the incongruent condition, $t(21) = 7.06$, $p < .001$, $d = 1.54$, and a facilitation in the congruent condition, $t(21) = 5.56$, $p < .001$, $d = 1.21$.

Are the effects constant throughout the test sessions?

In most Stroop paradigms, including the present experiment, there is the same number of congruent and incongruent items in order to comply with standard methodological requirements. However, by construction, the number of different incongruent items exceeds the number of different congruent items. As a consequence, congruent items are more often repeated than incongruent items, hence making it possible to learn from the sequence of items (see Melara & Algom, 2003). In the present experiment, the large number of different congruent items (i.e., 13) makes learning from the test somewhat unlikely. In addition, if some learning occurred during the experiment, this should be true for both musicians and nonmusicians. Therefore, the performance of nonmusicians should have departed from chance, which is not the case.

However, it remains possible that congruity effects increased throughout sessions for musicians, or still decreased, for instance, due to the gradual emergence of strategic factors. To address this issue, for each task, we divided the whole session ($N = 234$) into three blocks of equal length ($N = 78$), and the effect of congruity was computed for each block as the mean differences between scores for incongruent and congruent trials. There was no reliable change across sessions, whatever the task (note naming or word reading) and the dependent variable (errors or RTs) were. For the errors, an ANOVA with task (note naming, word reading) and block (1, 2, 3) as within-subject factors returned a main effect of task (see the next section for further analyses), but no main effect of block, $F(2, 42) = 0.45$, $p = .642$, and no Task × Block interaction, $F(2, 42) = 0.16$, $p = .851$. For RTs, the same analysis also gave a main effect of task (see the next section for further analyses), but no main effect of block, $F(2, 42) = 0.31$, $p = .738$, and no Task × Block interaction, $F(2, 42) = 1.44$, $p = .248$. Overall, these results indicate that the effects observed in musicians were a stable reflection of the expertise musicians have gained from their musical training in everyday life.

Comparing MSE and RMSE

An ANOVA on musicians' errors with condition (congruent, incongruent, neutral) and task (note naming, word reading) as within-subject variables revealed significant main effects of condition, $F(2, 42) = 27.24$, $p < .001$, $\eta_p^2 = .565$, and task, $F(1, 21) = 18.57$, $p < .001$, $\eta_p^2 = .469$, and a significant Condition × Task interaction, $F(2, 42) = 13.86$, $p < .001$, $\eta_p^2 = .398$. The congruity effect, as assessed by the difference between the incongruent and the congruent condition, was also stronger for the note-naming task ($M = 6.24\%$, $SD = 4.85$) than for the word-reading task ($M = 1.83\%$, $SD = 3.74$), $t(21) = 3.86$, $p = .001$, $d = 0.84$.

The analyses ran on musicians' RTs returned the same results, with significant effects of condition, $F(2, 42) = 68.39$, $p < .001$, $\eta_p^2 = .765$, and task, $F(1, 21) = 26.93$, $p < .001$, $\eta_p^2 = .562$, and a significant Condition × Task interaction, $F(2, 42) = 13.98$, $p < .001$, $\eta_p^2 = .40$. The congruity effect was much stronger for the note-naming task than for the word-reading task (106.30 ms vs. 31.36 ms, respectively), $t(21) = 4.47$, $p < .001$, $d = 0.98$. To sum up, the RMSE exceeded the MSE with both accuracy and RT measures.

Individual differences

In Cohen et al.'s (1990) model, the pattern of interference depends on a single parameter: the relative strength of the two competing pathways. If one postulates that RTs in the postexperimental ability tests provide a measure of strength (see below for a discussion), the relative strength of the word-reading and note-naming pathways for a given musician can be given by the difference of RTs in the two tasks (as mentioned above, error rates were negligible in the ability tests). Although naming notes took longer than reading for all musicians, the difference largely differed across musicians (range: from 17.61 ms to 230 ms). A participant P1 whose difference between note naming and word reading is larger than that for a participant P2 can be taken as having a stronger imbalance between musical and reading abilities than a participant P2. Is this difference actually predictive of the pattern of interference?

The response is clearly positive. There was a negative correlation between the relative strength of the two pathways and the amount of interference in the word-reading task, $r(20) = -.524$, $p = .012$, and there was a positive correlation between the relative strength of the two pathways and the amount of interference in the note-naming task, $r(20) = .428$, $p = .047$. To provide a more complete picture of the resulting pattern, participants were divided into three (roughly equal) groups along the relative strength dimension, with 7, 8, and 7 participants in each group. At one extreme of the resulting classification, the mean difference in speed between the two pathways was moderate (range: 17.61–76.58 ms). Participants exhibiting this pattern will be called *more balanced*. At the other extreme, the difference in strength between the two pathways is maximal (i.e., note naming is much longer than word reading, range: 181.06–230.00 ms). Participants exhibiting this pattern will be called *less balanced*. The remaining participants (*intermediate*) are in between.

Figure 5 shows the pattern of interference for each group. Overall, an ANOVA on the amount of interference with group as a between-subject variable (more balanced, intermediate, less balanced) and task (word reading, note naming) as a within-subject variable revealed no main effect of group, $F(2, 19) = 0.28$, $p = .756$, a significant effect of task, $F(1, 19) = 24.60$, $p < .001$, $\eta_p^2 = .564$, and, crucially, a significant Group × Task interaction, $F(2, 19) = 3.72$, $p = .043$, $\eta_p^2 = .281$. For the more-balanced group, the RMSE and the MSE were both different from zero, $t(6) = 2.75$, $p = .033$, $d = 1.24$, and $t(6) = 3.03$, $p = .023$, $d = 1.12$, respectively, and their amplitude did not significantly differ, $t(6) = 0.45$, $p = .672$. By contrast, for the less balanced group, the RMSE substantially exceeded the MSE, $t(6) = 6.68$, $p < .001$, $d = 2.73$, which was no longer significant, $t(6) = 1.76$, $p = .128$. For the intermediate group, the RMSE significantly exceeded the MSE, $t(7) = 5.18$, $p = .001$,

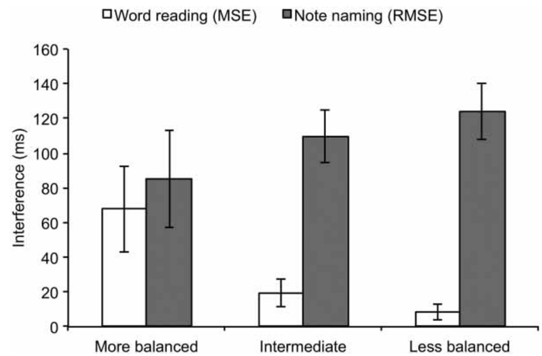

Figure 5. *Amount of interference (reaction times, RTs, in incongruent condition minus RTs in congruent condition) in the word-reading and note-naming tasks for three groups of musicians, sorted according to their performance in independent tests of reading and note-naming ability. MSE = musical Stroop effect; RMSE = reverse MSE. More balanced group: Note naming was only slightly longer than word reading. Less balanced group: Note naming was largely slower than word reading. Intermediate group: in between. Error bars indicate standard errors.*

$d = 1.96$, but the MSE was still reliable, $t(7) = 2.45$, $p = .044$, $d = 0.93$.

The same pattern was observed when the amount of interference was assessed through the rate of errors. As for RT measures, there was a negative correlation between the relative strength and the amount of interference in the word-reading task and a positive correlation between the relative strength and the amount of interference in the note-naming task. However, presumably due to the small number of errors, the correlations did not reach significance, $r(20) = -.321$, $p = .146$, and $r(20) = .347$, $p = .114$, respectively.

Discussion

Grégoire et al. (2013) reported a musical Stroop effect (MSE), whereby musicians showed slightly longer times for reading the name of a note printed in an incongruent position on a staff than for reading the same note name in a congruent position. The first contribution of the reported experiment was to confirm and strengthen this evidence, by showing a much larger interference of reading by note naming in a simplified paradigm. The second and main contribution of the experiment was to demonstrate that the same participants also exhibited interference of note naming by reading. Naming a note took longer when the note name written inside the note was incongruent with the note location on the staff than when it was congruent, giving evidence of a reverse MSE (RMSE). Incongruent patterns also generated significant increases of errors for both the MSE and the RMSE.

As pointed out in the introduction, reports of a reverse effect are not infrequent. However, the reverse effect is obtained in especially designed conditions, including the size and the orientation of the words (e.g., Palef & Olson, 1975), the level of training on a specific dimension (MacLeod & Dunbar, 1988), and the selection of participants (Akiva-Kabiri & Henik, 2012). The crucial point is that, in most cases, these conditions turned out to be inappropriate for the regular effect. Admittedly, a few studies (Melara & Mounts, 1993; Palef & Olson, 1975) have reported bidirectional effects without varying the conditions, but these effects were described as small and fleeting and could be due, for instance, to the mixture of unidirectional effects going in opposite directions for different participants (Cohen et al., 1990). As a consequence, the bulk of the evidence from the earlier literature is consistent with the conclusion that interference operates in a unidirectional way, which may alternate as a function of conditions or participants.

Our results are clearly at odds with this conclusion. The MSE and the RMSE were obtained in the very same conditions, which were chosen to match as well as possible the usual conditions of reading and note naming in real word settings. Moreover, 20 out of the 22 musicians showed the two effects on RTs (at a numerical level), excluding the possibility that bidirectional effects emerged from an artefact due to group statistics. Likewise, analyses of Vincent curves gave evidence that both the MSE and the RMSE were significant across the full range of latencies of the responses, even though the effects were stronger for the slower responses.[6] Finally, the idea that getting bidirectional effects could be dependent on the fact that all our participants were precisely at some optimal level of automatization in the two tasks (e.g., as in Macleod & Dunbar, 1988) is quite implausible: The selection criterion for our musicians was sufficiently loose (a minimum of 5 years of academic teaching) to allow the recruitment of participants largely differing on their level of musical proficiency. To conclude, the present study reports the strongest evidence to date for mutual interference between two competing processes.

The remainder of this paper first examines the implications of these results for Cohen et al.'s (1990) model, the main principles of which were described in the introduction. Then we address the question of why the results collected with note naming in this paper differ so strikingly from the results observed with colour naming or picture naming in the earlier literature.

Cohen et al.'s (1990) model

We pointed out above that Cohen et al. (1990) acknowledged that their difficulty in accounting for the mutual interference reported by MacLeod and Dunbar (1988) was a shortcoming of their model. However, insofar as mutual interference was only observed in a specific condition (at an intermediary level of training) by MacLeod and Dunbar, and looked as a nearly anecdotal phenomenon in the Stroop literature, the challenge seemed of limited importance. On the face on it, our results change the landscape and drastically increase the charges against the model.

However, we would like to argue that the problem raised by the observation of mutual interference is linked to the specific connectionist implementation and that, paradoxically, our data provide a striking support for the underlying psychological principles of the model. As noted in the introduction, Cohen et al. (1990) postulated

[6] Some proportionality between the size of a difference and the raw values on which this difference is computed is a common observation. In addition, the effect may be strengthened by the fact that slow responses may be less automatized and, hence, following Cohen et al.'s (1990) model, more receptive to interference than fast responses.

that the strength of a pathway both increases its ability to produce interference and reduces its susceptibility to interference. In the connectionist structure, this property generates mutual exclusivity of interference when two processes compete, because the stronger process both interferes on the weaker process and is immune to the interference the weaker process could produce. An effect or its reverse is obtained as a function of which of the two processes is the stronger. This implication does not seem to be a principled constraint, however. Rather than reasoning as if a change in the relative strength of the processes resulted in an all-or-none inversion in the direction of interference, the consequence could be conceived of as a gradual modification as well. The Cohen et al. claim, according to which it is "the relative strength, compared with a competing pathway [that is] important in determining whether a process will produce or be subject to interference in a Stroop-like task" (Cohen et al., 1990, p. 348) could thus be rephrased into a softer version: The relative strength compared with a competing pathway would be important in determining the relative propensity of the stronger process to interfere with and be interfered by another, weaker process, the final pattern taking the form of a trade-off between the amount of interference affecting each pathway (see Figure 2, for a schematic representation of the suggested modification to Cohen et al.'s model).

This softened version accounts for the whole pattern of results. The strength of word reading, as assessed by the speed of processing in the ability test, was considerably larger than the strength of note naming in musician participants. In keeping with the reasoning above, the stronger process, namely reading, elicited much larger interference on note naming than the reverse. But the most striking evidence came from the analyses of individual differences. The extent to which the strength of word reading exceeded the strength of note naming in musicians correlated negatively with the interference in word reading (MSE) and positively with the interference in note naming (RMSE). As shown in Figure 5, the subgroup of musicians who exhibited the best balanced performances in the naming test and the reading test also exhibited the best balanced amount of mutual interference, with the MSE being statistically indistinguishable from the RMSE. At the other end, the subgroup of musicians who exhibited the strongest imbalance in the naming test and the reading test also exhibited the strongest asymmetry of the effects, with the MSE being ultimately not statistically different from chance (note that these analyses were performed on very small samples of participants, which implies that statistical nonsignificance has only an indicative value). To sum up, the relative strength of the two competing pathways determined the pattern of interference, exactly as (a somewhat softened version of) Cohen et al.'s model would predict.

The analyses above take for granted that Cohen et al.'s concept of strength may be estimated by the speed of processing, and it could be argued that this postulate is questionable. Cohen et al. mentioned repeatedly that speed must be considered jointly with the amount of practice and the pattern of interference to assess strength. This argument calls for several comments. (a) Assuming that the duration of practice in a natural environment could be estimated with some confidence, directly comparing these durations between reading and note naming would be nearly meaningless. Indeed, the level of automaticity resulting from a same amount of practice obviously depends on the processes involved in the practised task. For instance, Logan and Klapp (1991) have shown that genuine properties of automaticity can emerge after less than 15 min of training on a simple and consistent alphabet-arithmetic task, an amount of practice that is not in the order of magnitude required to master complex activities such as reading. (b) Exploiting the pattern of interference to assess the strength of the pathways would be obviously circular in our case, given that we are trying to exploit the notion of strength to account for this pattern. (c) Although Cohen et al. took care to distinguish *strength* and *speed*, they also noted, for instance, that the smaller effects of words on picture naming than on colour naming "does not seem to be due to a difference in strength between picture naming and colour naming, because both have comparable reaction times in the control condition" (Cohen et al., 1990, their Footnote 18). This suggests that approximating the strength of a pathway through speed measures is not such a strong circumvention of Cohen et al.'s framework. (d) 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, but this does not entail discrediting speed as a useful correlate of automaticity, at least under specific conditions.

Finally, our assessment of Cohen et al.'s (1990) concept of *strength* through speed measures finds support in a last, more complex, but also more fundamental argument. The main problem of using speediness to measure strength is the same as that of using the amount of practice: All depends on the processes involved in the task, hence making between-task comparisons hazardous. In this regard, the claim above that reading is stronger than note naming because it is faster is admittedly questionable. But it is worth stressing that the part of evidence stemming from the analyses of individual differences does *not* face the same objection. Only within-task comparisons are exploited in these analyses. The only prerequisite is that *for a given task*, speed measures allow comparing the level of strength of the underlying processes between different participants. For instance, a participant P1 who reads notes faster than a participant P2 is taken as having a stronger pathway for note naming than P2, and likewise for reading. Now, a participant P1 for whom the difference between note naming and word reading is larger than that for a participant P2 can be taken as having a stronger asymmetry between musical and reading abilities than participant P2. This line of reasoning does not require that the absolute value of the difference between the speed of word reading and note naming makes sense.

To conclude, our findings have contrasted implications for Cohen et al.'s (1990) model. On the one hand, the observation of mutual interference for nearly all musicians despite their unequal ability to name notes runs clearly against the connectionist implementation of the model and presumably implies deep structural changes. On the other hand, however, our data provide a powerful support to the model intuition that the strength of a pathway both increases its ability to produce interference and reduces its susceptibility to interference, and that the final pattern of interference between two competing processes depends on their relative strength.

What is special about note naming?

So far, our analysis has focused on the ability of Cohen et al.'s (1990) model to account for the possibility of mutual interference. However, our results also raise a related, but different question: Why did these results depart from the data collected in earlier studies? In particular, why did note naming turn out to interfere with reading, whereas neither colour naming nor picture naming is able to do so without a strong perceptual degradation of the printed materials?

Cohen et al.'s (1990) model does not offer a clear response to this question. The model suggests that relative strength could be responsible, with note naming in musicians having greater strength than colour or picture naming. However, to avoid circularity, this interpretation should be validated through independent measures of strength, and this issue remains unsettled. Comparing the level of practice in natural settings between note naming on the one hand and colour naming or picture naming on the other hand makes little sense, for reasons discussed above. Moreover, measuring strength through the speed of responses does not support the hypothesis: RTs in the test of musical reading ability were not shorter and were in fact numerically longer (mean RT = 742 ms in our experiment) than those commonly reported for colour naming and picture naming, both of which are comparable (Cohen et al. estimate the mean RT to name colour or picture to be approximately 650 ms, see their Footnote 18). Note that these chronometric data also lead to disregard models relying directly on the relative speed of processing (e.g., Logan, 1980), as a possible explanation for the specificity of note naming with regard to colour or picture naming.

Melara and Algom (2003; see also Melara & Mounts, 1993) have questioned the relevance of a direct comparison between the strength (Cohen et al., 1990) or the speed (Logan, 1980) of the competing processes, as classically evaluated by the strength or speed of processing of individual stimulus attributes. They suggest rather that the relevant variable is the perceptual discriminability of the stimuli as assessed within each dimension. For each dimension, the level of discriminability is measured in the experimental context, by the speed or the accuracy of identification of two values alternating randomly along this dimension from trial to trial. The direction of the interference would depend on the relative discriminability of the values for the two competing dimensions, with the more discriminable dimension disrupting classification of the less discriminable dimension. Although these ideas have received some experimental support, their adequacy for accounting for our data is questionable. For our concern, at least two conditions should be fulfilled. The first is that the discriminability of musical notes (for musicians) should be much better than the discriminability of the colours or pictures used in standard Stroop tasks. Although further empirical studies would be needed at this point, we see no a priori reason that would lend support to this conjecture. Second and more importantly, one needs to posit that once the condition of matched discriminability is reached (or at least approximated), a sizeable mutual interference follows. The Melara and Algom (2003) model does not make this prediction, at least explicitly. At an empirical level, when discriminabilities were matched, Melara and Mounts (1993) observed "small interactive effect being reduced or eliminated through practice" (p. 627). Overall, irrespective of the interest of the Melara and collaborators' framework, it looks unlikely that it is appropriate for explaining our results.

At first glance, WEAVER++, the model proposed by Roelofs (2003, see also Roelofs, 1997, 2005), is more promising. Indeed, WEAVER++ successfully accounts for the mutual interference observed in MacLeod and Dunbar (1998) at an intermediate stage of practice. The predictions of WEAVER++ are grounded on the prior knowledge of the task architectures, gained by independent evidence coming from behavioural and neuropsychological investigations. The asymmetry between reading and colour naming is explained within the theory of Levelt, Roelofs, and Meyer (1999), according to which a process of conceptual identification would precede articulation in colour naming, whereas the order of these two steps of processing would be reversed for reading aloud. As a consequence, colour would require an extra processing step before reaching articulation compared to word.

The Roelofs model predicts mutual interference whenever two tasks involve the same architecture. The explanation for mutual interference in MacLeod and Dunbar (1998) relies on the fact that none of the tasks used in this study—colour naming and shape naming—involved the reading architecture. Given that the reading architecture is necessarily involved as one of the two dimensions in our procedure, the reason why WEAVER++ achieves to account for mutual interference in MacLeod and Dunbar cannot be transposed to our experiment. The only solution to account for our results in the Roelofs model would consist in attributing to note naming the same architecture as that for reading, or at least an architecture that would be closer to reading than to colour naming or picture naming. Our lack of background knowledge on the architecture of the note-naming pathway makes further speculations hazardous, but we submit nevertheless that this condition seems to be hard to meet. Indeed, on the face of it, note naming is nothing else but an instance of picture naming (note we are dealing here only with note naming such as that involved in our experiment, and not with the full array of processes involved in music reading). A possibility would be that with very extensive training, the extra processing step involved in note naming turns out to be by-passed. Further studies on the psychological processes involved in note naming in expert musicians are needed before firmer conclusions may be drawn.

Another popular model of the Stroop effect is the translational account (e.g., Durgin, 2003; Virzi & Egeth, 1985). This model is aimed at accounting for the asymmetry commonly reported in the literature between reading and any other competing dimension. The underpinning idea is that interference arises only if responding to the target implies a translation from one code to another code. For instance, naming a colour implies for the colour code to be converted into a verbal code, and this translation makes the vocal response vulnerable to interference. By contrast, the written word is already in the verbal code serving for the response, and, hence, reading aloud is immune to any interference.

As such, the translational model is obviously unable to account for the part of our results evidencing interference on reading, and in this regard, the reported evidence for an MSE sounds like a strong case against the model (for other contradictory data, see Blais & Besner, 2007). Could the translational model be completed or modified to account for mutual interference nevertheless? The only solution we see as compatible with the model would be to add the postulate that two competing processes involving no translation could exhibit mutual interference. To work in the present situation, this would need to consider that note naming implies no translation, hence differing in this regard from both colour naming and picture naming. Unfortunately, naming a note seemingly implies a translation from a pictorial code to a verbal code, in the very same way as naming a picture.

To conclude, our examination of different concepts involved in the current Stroop models to account for the direction of interference—strength, speed of processing, perceptual discriminability, task architecture, or still code translation—leaves us rather somewhat pessimistic on their propensity to explain the particularity of note naming with regard to the more common dimensions competing with reading in Stroop paradigms.

Reconsidering an old idea

We suggest that the discrepancy between our and earlier results could be explained in reference to the early observation of Peterson, Lanier, and Walker (1925; cited in MacLeod, 1991; Peterson was the supervisor of John Ridley Stroop's thesis). Peterson et al. wrote:

> To the written words "red", "blue", "green" etc., the subjects have as a rule given in the past but the one response of pronouncing (vocally or subvocally) the names of these colors; whereas on seeing the colors red, blue, green, etc. they have responded in many different ways, as grasping and eating, handling, perceiving and admiring, etc. In the case of the words, then, but one specific response-habit has become associated with each word, while in the case of the colors themselves a variety of response tendencies has been developed. (p. 281)

Our claim is that note naming is much closer to reading than to colour or picture naming on that dimension. As a rule, a musician gives a response of pronouncing vocally (e.g., at the beginning practice) or subvocally (e.g., when playing an instrument later in practice) the name of the note when exposed to a note on a staff.

To our knowledge, this idea has never been articulated as a full-blown model in the Stroop literature, although similar suggestions have been done on occasions (e.g., Cohen et al., 1990, their Footnote 7). However, there is at least another research domain in which a related idea has been heavily developed. We allude here to the research on associative learning, to which Peterson et al. (1925) refer. Since Rescorla (1968), the associative learning tradition, extended by recent studies on statistical learning (see Perruchet & Poulin-Charronnat, 2012, for linking the two domains together), emphasizes the pervasive role of the *contingency* between events in the strengthening of associative relations. The measures of contingency differ from the simple frequency of co-occurrences: The level of contingency indicates how an event is *predictive* of another event. Note that Shiffrin and Schneider (1977, see also Schneider & Shiffrin, 1977) have provided impressive evidence for the role of contingency in the formation of automatisms, although they did not used the term of contingency. They refer to the *consistent mapping* of stimulus–response pairs, by contrast to varied mapping, to describe a condition that could also be described as a perfect contingency between events.

To reframe Peterson et al.'s (1925) intuition in contemporary terminology, one could say that, in real-world life, there is a low level of contingency between seeing a colour and producing the name of the colour. The main reason is not that producing the name of a colour in response to a coloured patch is infrequent: After all, one could argue that children are doing that intensively when learning the colours. The main reason is that colours are continually present in our visual field without their names being evoked, overtly or covertly. The very same comment could be made about picture naming: We are continually faced with objects that, in an overwhelming proportion of situations, we are not naming. By contrast, the contingency between seeing a written word and producing, vocally or subvocally the word is certainly considerably stronger. What about note naming in musicians on this dimension? Although highly speculative, the idea may be defended that the contingency between perceiving a note on a staff and generating the name of this note is at least equal to, if not stronger than, the stimulus–response contingency in word reading. Indeed, even for musicians, a note is virtually never displayed in the environment out of the context in which it ought to be named.

If one endorses this framework, the conclusions stemming from the standard paradigms—the colour–word and picture–word tasks—would be limited to the case where the process competing with word reading has evolved in conditions that were not conducive to automatization. Using note naming in musicians suggests a new conjecture: When reading competes with another process that is highly automatized due to the strong stimulus–response contingency, getting mutual interference could be the rule.

Original manuscript received 8 April 2013
Accepted revision received 25 November 2013
First published online 2 April 2014

REFERENCES

Akiva-Kabiri, L., & Henik, A. (2012). A unique asymmetrical stroop effect in absolute pitch possessors. *Experimental Psychology*, *59*(5), 272–278.

Blais, C., & Besner, D. (2006). Reverse stroop effects with untranslated responses. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1345–1353.

Blais, C., & Besner, D. (2007). A reverse Stroop effect without translation or reading difficulty. *Psychonomic Bulletin & Review*, *14*(3), 466–469.

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.

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*(2), 470–484.

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.

Durgin, F. H. (2003). Translation and competition among internal representations in a reverse Stroop effect. *Perception & Psychophysics*, *65*(3), 367–378.

Ellis, N. R., Woodley-Zanthos, P., Dulaney, C. L., & Palmer, R. L. (1989). Automatic-effortful processing and cognitive inertia in persons with mental retardation. *American Journal on Mental Retardation*, *93*(4), 412–423.

Flowers, J. H., Warner, J. L., & Polansky, M. L. (1979). Response and encoding factors in "ignoring" irrelevant information. *Memory & Cognition*, *7*(2), 86–94.

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.

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 640–654.

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The musical Stroop effect: Opening a new avenue to research on automatisms. *Experimental Psychology*, *60*(4), 269–278.

Grégoire, L., Perrcuhet, P., & Poulin-Charronnat, B. (2014). Is the musical Stroop effect able to keep its promises? A reply to Akiva-Kabirai and Henik (2014), Gast (2014), Moeller and Frings (2014), and Zakay (2014). *Experimental Psychology*, *61*(1), 80–83.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38.

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology*, *12*(4), 523–553.

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is Extended practice necessary to produce automaticity?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(2), 179–195.

Lowe, D. G., & Mitterer, J. O. (1982). Selective and divided attention in a Stroop task. *Canadian Journal of Psychology*, *36*(4), 684–700.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203.

MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition*, *26*(2), 201–211.

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.

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*(10), 383–391.

Melara, R. D., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review*, *110*(3), 422–471.

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.

Palef, S. R., & Olson, D. R. (1975). Spatial and verbal rivalry in a Stroop-like task. *Canadian Journal of Psychology*, *29*(3), 201–209.

Perruchet, P., & Poulin-Charronnat, B. (2012). Word segmentation: Trading the (new, but poor) concept of statistical computation for the (old, but richer) associative approach. In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 119–143). Berlin: De Gruyter Mouton.

Peterson, J., Lanier, L. H., & Walker, H. M. (1925). Comparisons of white and negro children in certain ingenuity and speed tests. *Journal of Comparative Psychology*, *5*, 271–283.

Protopapas, A., Archonti, A., & Skaloumbakas, C. (2007). Reading ability is negatively related to Stroop interference. *Cognitive Psychology*, *54*(3), 251–282.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*, 446–461.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology*, *66*, 1–5.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*, 249–284.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review*, *110*(1), 88–125.

Roelofs, A. (2005). The visual-auditory color-word stroop asymmetry and its time course. *Memory & Cognition*, *33*(8), 1325–1336.

Roelofs, A. (2008). Tracing attention and the activation flow in spoken word planning using eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 353–368.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*(2), 127–190.

Stewart, L., Walsh, V., & Frith, U. (2004). Reading music modifies spatial mappings in pianists. *Perception & Psychophysics*, *66*(2), 183–195.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662.

Verdonschot, R. G., La Heij, W., & Schiller, N. O. (2010). Semantic context effects when naming Japanese kanji, but not Chinese hànzì. *Cognition*, *115*, 512–518.

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory & Cognition*, *13*(4), 304–319.

Zakay, D., & Glicksohn, J. (1985). Stimulus congruity and S-R compatibility as determinants of interference in a Stroop-like task. *Canadian Journal of Psychology*, *39*(3), 414–423.