Research Article

The Musical Stroop Effect

Opening a New Avenue to Research on Automatisms

Laurent Grégoire, Pierre Perruchet, and Bénédicte Poulin-Charronnat

LEAD-CNRS, University of Bourgogne, Dijon, France

Abstract. The usual color-word Stroop task, as well as most other Stroop-like paradigms, has provided invaluable information on the automaticity of word reading. However, investigating automaticity through reading alone has inherent limitations. This study explored whether a Stroop-like effect could be obtained by replacing word reading with note naming in musicians. Note naming shares with word reading the crucial advantage of being intensively practiced over years by musicians, hence allowing to investigate levels of automatism that are out of reach of laboratory settings. But the situation provides much greater flexibility in manipulating practice. For instance, even though training in musical notation is often conducted in parallel with the acquisition of literacy skills during childhood, many exceptions make that it can be easily decoupled from age. Supporting the possibility of exploiting note naming as a new tool for investigating automatisms, musicians asked to process note names written inside note pictures in incongruent positions on a staff were significantly slowed down in both a go/no-go task (Experiment 1) and a verbal task (Experiment 2) with regard to a condition in which note names were printed inside note pictures in congruent positions.

Keywords: Automatism, Stroop effect, interference, musical expertise, note naming

Stroop's (1935) classic article is certainly one of the most influential papers on cognition. The effect reported in this article is well known: Color naming is slowed down by an incongruent color word. The huge interest for this phenomenon stems from the fact that it allows exploring one of the major properties of automatisms, namely the fact that after extensive practice with a consistent task, the processes involved in this task tend to generate interference with other tasks in which they are in principle irrelevant.

A number of variants to the standard color-word version have been created, which were essentially devised to provide an alternative to the color dimension. In the picture-word paradigm, for instance, a word (e.g., "hand") is printed inside a congruent or an incongruent picture (e.g., a hand or a foot, respectively; e.g., Lupker, 1979), while in a spatial version of the Stroop task, the word "below" is displayed above a fixation point, or conversely (e.g., Palef & Olson, 1975). In both cases and many other Stroop-like paradigms (although not all; exceptions will be dealt with in the General Discussion), interference is generated by word reading.

Investigating word reading is of obvious interest, given the crucial importance of the issue in daily behavior. But focusing on reading is endowed with its own shortcomings. First, any generalization of conclusions drawn from a single experimental arrangement remains a matter of speculation. More importantly, reading offers quite limited opportunity to manipulate certain variables that would seem of primary interest when studying interference and automaticity. As claimed by MacLeod (1991), "practice may turn out to be one of the most effective manipulations for disentangling theories of the Stroop effect" (p. 182). Now, manipulating reading practice is severely restrained due to obvious practical and ethical constraints. For instance, the only possible approach to assess the amount of reading practice consists in using academic level or reading skill as an independent variable (e.g., Catling, Dent, Johnston, & Balding, 2010; Protopapas, Archonti, & Skaloumbakas, 2007). Unfortunately, reading ability is acquired within an age span where an overwhelming amount of other cognitive changes occurs, and the observed performance evolution may be due to multiple factors other than reading practice, if only the correct understanding of instructions is shown by the youngest children.

In Search for an Alternative Paradigm

In this paper, we propose to trade word reading for note naming in musicians, hence analyzing the possibility of observing what is coined herein as the *musical Stroop effect*. 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. A musical Stroop effect would be revealed by the impaired processing of the printed note name in incongruent conditions with regard to congruent conditions, an effect that would attest to the interference generated by note naming in musicians.

The musical Stroop would share most advantages of the classical color-word version to investigate automatisms. The most obvious is that note naming is an activity that is intensively practiced over years by musician experts, and relying on naturally occurring practice allows attaining a level of practice that is out of reach of laboratory settings. Moreover, note naming shares with word reading the nice

DOI: 10.1027/1618-3169/a000197

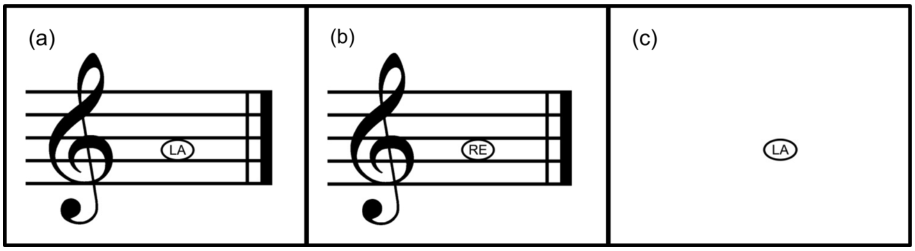

Figure 1. Examples of the congruent-context (a), the incongruent-context (b), and the out-of-context conditions (c). Note that in the musical French notation (and several other countries such as Italy and Spain), note names are DO, RE, MI, FA, SOL, LA, SI, instead of the first letters of the alphabet.

property of being a component of more complex activities, the automaticity of which, far from generating anecdotic glitches, is required to ensure the successful expression of the whole behavior. The automaticity of note naming is necessary to allocate musicians' attention to higher integrative processes devoted to analyzing chords and melodic lines of the musical work and to ensure motor control for instrument playing, as the automaticity of identification of single words is needed to free our minds to deal with higher-order aspects of the task such as comprehension and metacognitive functions.

However, as an experimental means of investigation, the musical Stroop effect would avoid the shortcomings pointed out above, because musical formation provides much greater flexibility in manipulating practice than literacy learning. Control over the practice level is tighter, because music notes are not ubiquitously present, as words are, in our daily environment. Moreover, even though musical training is often conducted in parallel with the acquisition of basic literacy skills during childhood, there are a large number of exceptions, which make that practice can be easily decoupled from age. For instance, instead of running experiments with children aged from 6 to 11 years to investigate the first years of practice on word reading, exploiting the musical Stroop effect makes it possible to run experiments with 12-year-old children having from 1 to 6 years of musical practice. Conversely, it becomes possible to explore whether automatism formation interacts with age, by comparing the effect of extensive practice at various ages, even including elderly people (exploiting the fact that a proportion of newly retired persons start musical training). Among a host of other possibilities offered to experimental exploration, and without any possible analog in reading, is the fact that a number of persons having gained a high level of expertise in music turn out to give up any practice for years.

Is There Previous Evidence for a Musical Stroop Effect?

Zakay and Glicksohn (1985) exposed pianists to musical notes or to the names of those notes, which were written at congruent or incongruent locations on a staff. Participants performed verbal and manual responses in succession, in counterbalanced order. For the verbal response, they were required to name the notes according to their position on the staff while ignoring the printed name, or alternatively, to read aloud the printed names of the notes while ignoring their position on the staff. The manual response involved the same conditions, except that participants were required to press the appropriate piano key. Zakay and Glicksohn reported an overall effect of congruency. However, this effect interacted with the mode of response, and Zakay and Glicksohn did not assess the effect of congruency separately for each response mode (verbal vs. manual). More importantly, Zakay and Glicksohn did not distinguish the effect of congruency due to the automaticity of word reading (as virtually any other Stroop-like tasks) from the effect of congruency due to musical abilities, which is at focus in the present study.

Zakay and Glicksohn, nevertheless, provided detailed response times for each condition (see their Table 1, p. 418) and asserted, on the basis of a post hoc Scheffé procedure, that a difference of 880 ms was required for any contrast to be significant at the .05 level.[1] Using this conservative benchmark, it appears that the musical Stroop effect did not reach significance: Incongruent conditions slowed down verbal responses, but to a far too limited extent (310 ms). The main effect of congruence was in fact driven by the conditions in which pianists had to respond through piano keys.[2] Although these early results raise some skepticism about the possibility of observing a

[1] This is the difference in RTs for a whole sequence of 10 homogeneous stimuli.
[2] These data are potentially relevant for research on motor skills. However, a number of factors might have artificially inflated, if not entirely producing the congruence effect. For instance, when pianists have to press, say, a ''SOL'' written on the ''LA'' line in incongruent conditions, their response may be delayed by the interference due to the irrepressibility of the response ''LA'' (a Stroop-like effect), but the delay may also be due to the fact that the word ''SOL'' does not designate a single key on the piano keyboard. Subjects have to select arbitrarily at least between the two SOL keys surrounding the last pressed key, which one to press, and this selection is presumably time consuming.

musical Stroop effect as defined above, it cannot be excluded that some features of the Zakay and Glicksohn's experiment, the description of which are postponed to the General Discussion, could have played down the effect measured with verbal responses.

The Present Study

The experiments below explored again the potential of using the automaticity of note naming in musicians instead of reading as a source of a Stroop-like effect. Our procedure is a conceptual replication of a subset of the conditions implemented in the Zakay and Glicksohn's (1985) study, with the addition of a number of controls and methodological improvements. The main experimental conditions (i.e., congruent and incongruent) were presented above (Figure 1a and b). In addition, words that were not names of notes were also printed inside the notes on some trials. In Experiment 1, both musicians and nonmusicians were told to press a key when the printed word was a note name, and to refrain any response when it was not. In Experiment 2, musicians and nonmusicians were instructed to read aloud printed words, while ignoring note positions. In both experiments, a musical Stroop effect should result in (1) longer response times in the incongruent than in the congruent condition for musicians and (2) an interaction of the congruity effect with musical expertise, attesting that the congruity effect is specific to musicians, and hence is likely to be a genuine consequence of musical practice.

A long-lasting debate in the Stroop literature is whether the Stroop effect, classically defined as the difference between performances in congruent and incongruent conditions, is mainly due to interference in the incongruent condition, facilitation in the congruent condition, or some mix of both effects. To explore this issue, a baseline situation was added, in which the note names were printed inside a picture of a note, but outside any contextual staff (Figure 1c).

Experiment 1

Both intuition and experimental data suggest that reading a word aloud may be less susceptible to interference than a more delayed response involving subsequent steps of processing. Earlier Stroop literature has shown that classifying the word as belonging to a predefined category (e.g., responding "animal" when the word is "frog" in a picture-word task, see Smith & Magee, 1980) increased the probability of interference with the irrelevant dimension. Independently, using a motor response instead of word reading (e.g., pressing color-labeled buttons in the color-word version, see Blais & Besner, 2006) is endowed with the same consequence. To increase our chance of getting an effect in the first experiment, these two methodological variants were combined. A go/no-go task was used (as Durgin, 2003), in which participants had to press a key only when the printed word was a note name. A selective slowdown in response time of musicians when the printed note name was incongruent with the note picture would attest to a musical Stroop effect.

Method

Participants

Twenty-eight volunteer undergraduate psychology students at the Université de Bourgogne received course credit for their participation. They were French native speakers and reported normal or corrected vision. Half of them had formal musical training and had played a musical instrument for at least 5 years (musicians: eight women, six men). They had an average of 13.93 (SD = 3.77) years of musical training. Other participants had never studied or practiced music (nonmusicians: nine women, five men).

Materials

For the main three experimental conditions, note names were printed inside note pictures, which could appear on each of the 13 possible positions going from C4 to A5. In the congruent- and incongruent-context conditions, the note pictures were presented in a treble staff (Figure 1a and b). For the congruent condition, the note name was congruent with the note position on the staff, 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.

In the out-of-context condition, the note pictures were presented without the staff, although for the sake of matching at best perceptual conditions, they appeared on the same 13 spatial locations, as if they were correctly positioned on a virtual staff (Figure 1c).

For the purpose of the go/no-go task, the to-be-read stimuli of two additional conditions were made up of words that were not names of notes. As the note names, they were six two-letter and one three-letter highly frequent French words (*CE, JE, TU, NI, TA, VU*, and *PAR*). These words were printed inside note pictures, and they were displayed at all 13 note locations either inside a staff (in-context condition) or without any surrounding staff (out-of-context condition).

Note names and non-note words appeared in standard uppercase font printed in black over a white background on a computer screen. The treble staff was 7.7 cm wide by 5.1 cm high. For each of the five conditions (congruent, incongruent, note names out-of-context, words in-context, and words out-of-context), the stimuli appeared six times on each of the 13 locations, leading to 78 trials per condition, and resulting in 390 trials (78 × 5) for the whole session.

Procedure

Participants were asked to press the space bar, as quickly as possible, when the printed word was a note name, and to refrain from responding when it was not. The next stimulus appeared after 1,200 ms if no response had been made.

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. A fixation cross displayed for 1 s at the center of the screen preceded the stimulus, which stayed on the screen until participant's response. The intertrial interval was 1 s. The 390 trials were pseudorandomly ordered for each participant, excluding immediate repetitions of note locations, note names, and non-note words. They were displayed as 10 blocks of 39 trials each with a self-paced break between blocks. After the session, which lasted about 20 min, the musicians filled out a questionnaire about their musical abilities. This questionnaire included information about the number of years of musical training, which is reported above, and a lot of other questions that were not exploited, such as the instrument that was played (if any), the amount of daily practice, and whether participants possessed absolute pitch.

Results and Discussion

Values beyond three standard deviations of the mean (1.03%) were removed from the data for each participant. Misses and false alarms represented 1.04% and 3.37% of the trials, respectively.

Mean Response Times (RTs) for hits are shown in Figure 2. RTs were significantly longer in incongruent- than in congruent-context condition for musicians, with a mean difference of 9.91 ms (SD = 11.56), $t(13) = 3.21$, $p = .007$, $d = .86$. There was no corresponding effect of congruity for nonmusicians, $t(13) = 0.66$, $p = .521$. To assess whether the congruity effect significantly differed between the two groups, an ANOVA was carried out with Congruity (congruent-context, incongruent-context) as a within-subject variable and Musical Expertise (musicians, nonmusicians) as a between-subject variable. There was no main effect of congruity, $F(1, 26) = 2.47$, $p = .128$, a significant effect of musical expertise, $F(1, 26) = 33.11$, $p < .001$, $\eta_p^2 = .560$, and crucially a significant Congruity × Musical Expertise interaction, $F(1, 26) = 6.64$, $p = .016$, $\eta_p^2 = .203$.

The out-of-context condition was devised to provide a baseline for assessing whether the Stroop effect observed in musicians was due to interference or facilitation. An ANOVA with Condition (congruent-context, incongruent-context, out-of-context) as a within-subject variable and Musical Expertise (musicians, nonmusicians) as a between-subject variable gave significant effects for condition, $F(2, 52) = 101.35$, $p < .001$, $\eta_p^2 = .796$, musical expertise, $F(1, 26) = 35.142$, $p < .001$, $\eta_p^2 = .575$, and Condition × Musical Expertise interaction, $F(2, 52) = 4.33$, $p = .018$, $\eta_p^2 = .143$. RTs of musicians were shorter in the out-of-context condition than in the incongruent-context condition, $t(13) = 10.90$, $p < .001$, $d = 2.91$, but surprisingly, they were also shorter than in the congruent-context condition, $t(13) = 9.42$, $p < .001$, $d = 2.52$. This pattern suggests that even the congruent condition generates interference, a result that has sometimes been reported in the color-word version of the Stroop task (e.g., Nealis, 1973; Schulz, 1979). However, it turns out that the same effects occurred for nonmusicians. RTs of nonmusicians were also shorter in the out-of-context condition than in both the incongruent-context condition, $t(13) = 7.52$, $p < .001$, $d = 2.00$, and the congruent-context condition, $t(13) = 6.20$, $p < .001$, $d = 1.66$, suggesting that musical expertise does not play any role in this effect. Another interpretation for shorter RTs in the out-of-context condition stems from the degree of perceptual complexity of each condition. Word reading would be easier when the note is displayed alone than when it appears inside a complex frame staff.

If one assumes that musicians and nonmusicians are both sensitive to perceptual complexity, then the differences between musicians and nonmusicians in the congruent- and incongruent-context conditions can be compared to the difference between musicians and nonmusicians in the out-of-context condition. In the out-of-context condition, musicians were 127.35 ms faster than nonmusicians. The corresponding value for the congruent-context condition was nearly identical, $M = 125.12$, $F(1, 26) = 0.15$, $p = .706$, whereas the difference for the incongruent-context condition was significantly reduced, $M = 112.81$, $F(1, 26) = 7.59$, $p = .011$, $\eta_p^2 = .226$. In other words, musicians and nonmusicians differed only in the incongruent condition once their difference in the out-of-context condition has been partialled out, which suggests that the musical Stroop effect is mainly due to interference in the incongruent condition.

To conclude, we obtained clear evidence for a musical Stroop effect using a go/no-go procedure: As anticipated, there was a reliable effect of congruity, and this effect was limited to musicians. On the other hand, data also revealed an unexpected main effect of musical expertise: In all conditions, musicians performed the go/no-go task considerably faster than nonmusicians. Recall that in the go/no-go task, participants had to decide whether the printed item was a note name or not. Although nonmusicians are familiar with note names, if only because they are part of many popular children's songs, it is very likely that the exhaustive set of note names is more accessible for musicians than for nonmusicians. A possibility is that a decision of categorical membership would require serial memory search in nonmusicians, while retrieval of the whole set of note names would be performed in parallel in musicians (Shiffrin & Schneider, 1977).

Experiment 2 aimed at addressing three remaining questions. First, as discussed above, the go/no-go task (a categorization task, which involves a motor mode of responding)

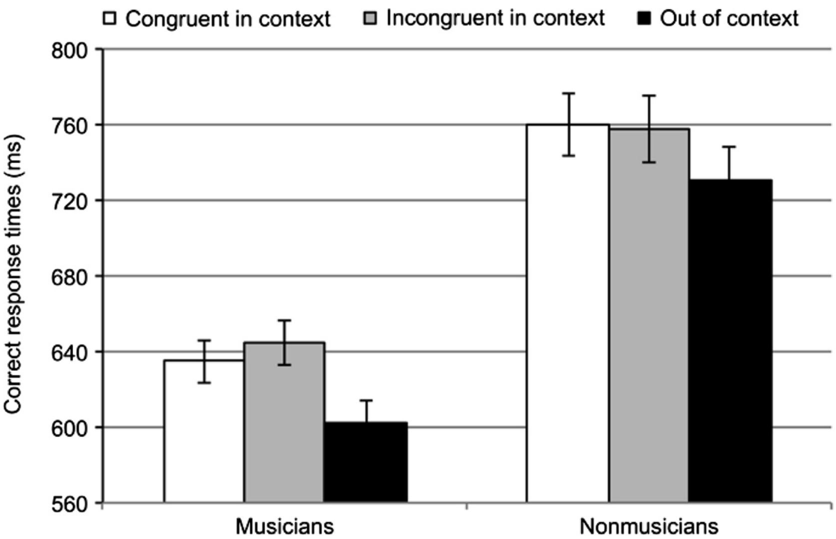

Figure 2. Correct response times as a function of context and musical expertise in Experiment 1. Error bars indicate standard errors.

cumulates two procedural changes known to make the required responses more susceptible to interference than reading aloud. Using this task does not invalidate or undermine our conclusion, but gives rise to our main question: Is a musical Stroop effect still present when the go/no-go task is replaced with more standard oral responses? Second, are the longer RTs observed in the in-context than in the out-of-context condition for both musicians and nonmusicians really due to the relative perceptual complexity of the respective displays? Third, is the unexpected overall slowness of nonmusicians due to the categorical membership decision required in the go/no-go task?

Experiment 2

Experiment 2 involved the very same five conditions as Experiment 1. However, participants had to read aloud the printed word (a note name or another word), while ignoring the position of the note. Reading times collected in the congruent and incongruent conditions should allow confirming evidence for a musical Stroop effect. Trials involving words that were not names of notes, which served only as foils in the go/no-go task of Experiment 1, were exploited in Experiment 2 to indicate whether the presence of the staff context is sufficient to slow down the responses. Because the notion of congruity is objectless for these items, any difference in RTs between in-context and out-of-context conditions would reveal the influence of a variable unrelated to musical information *per se*, presumably perceptual complexity. This information should allow assessing whether the RTs observed in the out-of-context condition involving notes as items can serve as a baseline for teasing apart inhibitory and facilitatory components of the musical Stroop effect. Finally, because a categorical membership decision is no longer involved, Experiment 2 should allow deciding whether the overall advantage of musicians over nonmusicians observed in Experiment 1 was task specific.

Method

Participants

Thirty-four new participants were recruited from the same population as in Experiment 1. Half of them had formal musical training and had played a musical instrument for at least five years (musicians: 11 women, 6 men). They had an average of 10.12 (SD = 3.86) years of musical training. Other participants had never studied or practiced music (nonmusicians: 12 women, 5 men).

Materials and Procedure

The materials and procedure were identical to Experiment 1, except that participants were now asked to read aloud the printed word, while ignoring the context. They were encouraged to respond as quickly and accurately as possible. The RTs were recorded by a voice key. During the session, the experimenter noted the error responses and the voice key dysfunctions.

Results and Discussion

Voice key dysfunctions led to exclude 2.82% of the data. Reading errors (0.02%) and RTs beyond three standard deviations of the mean (0.70%) were removed. Mean RTs for correct responses are shown in Figure 3.

RTs to Note Names

For musicians, RTs were again significantly longer in incongruent- than in congruent-context condition, with a mean difference of 8.69 ms (SD = 16.46), $t(16)$ = 2.18,

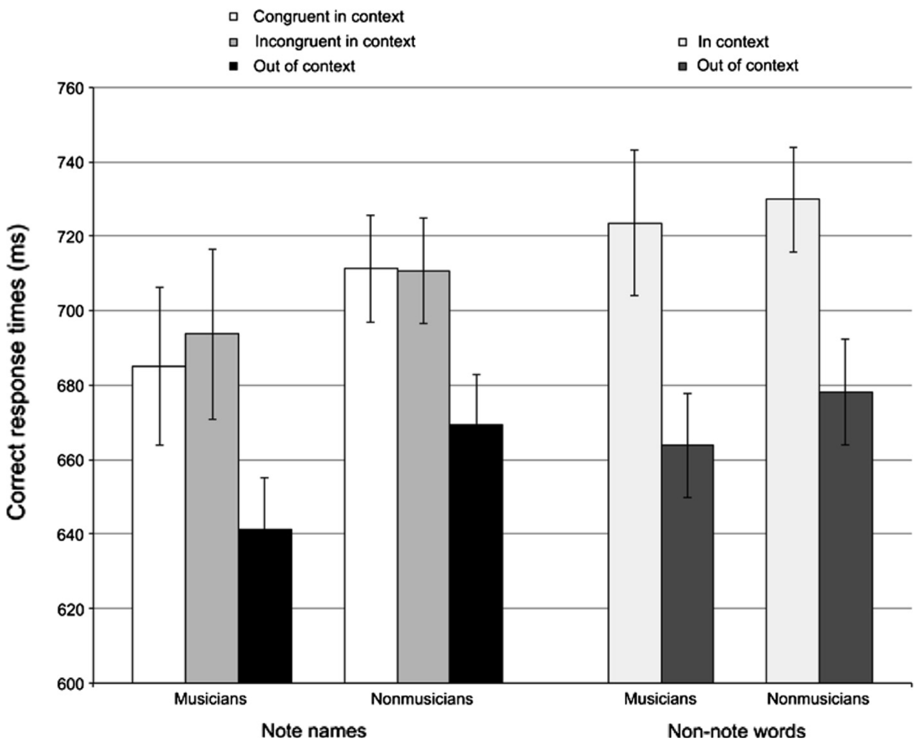

Figure 3. Correct response times as a function of context and musical expertise in Experiment 2. Error bars indicate standard errors.

$p = .045$, $d = .53$. There was no congruity effect for nonmusicians, $t(16) = 0.25$, $p = .803$. An ANOVA performed with Congruity (congruent-context, incongruent-context) as a within-subject variable and Musical Expertise (musicians, nonmusicians) as a between-subject variable revealed no main effect of musical expertise, $F(1, 32) = 0.75$, $p = .394$, and the congruity effect was only marginally significant, $F(1, 32) = 3.29$, $p = .079$, $\eta_p^2 = .093$. More importantly, there was a significant Congruity × Musical Expertise interaction, $F(1, 32) = 4.20$, $p = .049$, $\eta_p^2 = .116$. To sum up, we obtained clear evidence for a musical Stroop effect, with a congruity effect limited to musicians, hence replicating the results from Experiment 1.

Figure 3 also illustrates that reading note names was much faster in the out-of-context condition with regard to the congruent and the incongruent conditions for both musicians and nonmusicians, as in the go/no-go task used in Experiment 1. An ANOVA with Condition (congruent-context, incongruent-context, out-of-context) as a within-subject variable and Musical Expertise (musicians, nonmusicians) as a between-subject variable gave only a main effect of condition, $F(2, 64) = 60.88$, $p < .001$, $\eta_p^2 = .655$. Neither musical expertise effect, $F(1, 32) = 1.10$, $p = .302$, nor the Condition × Musical Expertise interaction, $F(1, 32) = 0.76$, $p = .472$, reached significance. Planned comparisons showed that RTs for the out-of-context condition were lower than both the incongruent, $F(1, 32) = 62.98$, $p < .001$, $\eta_p^2 = .663$, and the congruent conditions $F(1, 32) = 69.11$, $p < .001$, $\eta_p^2 = .684$.

RTs to Non-Note Words

Crucially, there was also a clear effect of context when non-note words were used instead of note names in the two additional conditions. An ANOVA was performed on the RTs for non-note words with Context (in-context, out-of-context) as a within-subject variable and Musical Expertise (musicians, nonmusicians) as a between-subject variable. There was a significant and very large effect of context, $F(1, 32) = 180.50$, $p < .001$, $\eta_p^2 = .849$, whereas neither the musical expertise effect, $F(1, 32) = 0.24$, $p = .625$, nor the Context × Musical Expertise interaction, $F(1, 32) = 0.93$, $p = .341$, was significant. These results support our account of the results from Experiment 1 in terms of perceptual complexity.

Facilitation or Interference?

As in Experiment 1, it remains possible to use the difference between musicians and nonmusicians in the out-of-context condition as a baseline for interpreting the differences between musicians and nonmusicians in the congruent- and incongruent-context conditions. In the out-of-context condition, musicians were 27.78 ms faster than nonmusicians. The difference for the congruent-context condition was nearly identical, $M = 26.27$, $F(1, 32) = 0.02$, $p = .885$. By contrast, the difference for the incongruent-context condition was numerically reduced ($M = 17.05$), although the reduction did not reach significance, $F(1, 32) = 0.82$, $p = .371$. This pattern is qualitatively similar to the pattern observed in Experiment 1, suggesting that congruence would elicit no beneficial effect in musicians, whereas incongruence would elicit a detrimental effect.

Finally, although the RTs of musicians were numerically shorter than the RTs of nonmusicians in all conditions, the analyses above show that the effect never reached significance. This suggests that the very strong effect observed in Experiment 1 was primarily due to the go/no-go task, in which the categorical membership decision was presumably very sensitive to the level of accessibility of the set of note names in the two groups of participants.

General Discussion

Musicians processed note names written inside note pictures in incongruent positions on a staff significantly slower in both a go/no-go task (Experiment 1) and a verbal task (Experiment 2) with regard to a condition in which note names were printed in congruent positions. Our results also showed that reading times of note names printed outside a staff cannot serve as an appropriate baseline to assess the proportion of the effect due to interference and facilitation. Indeed, inserting the note inside a staff slowed down reading times in Experiment 2, for both note names and nonnote words, and in both musicians and nonmusicians, hence suggesting that the staff increased the perceptual difficulty of the task independently of its informational value for musicians. Assuming that the effect of perceptual complexity was identical for musicians and nonmusicians, however, our data suggest that congruence elicited no beneficial effect in musicians, whereas incongruence elicited a detrimental effect. This overall prevalence of interference over facilitation has been reported in nearly all other versions of the Stroop effect (e.g., MacLeod, 1991).

This study therefore gives the first compelling evidence for a musical Stroop effect, defined as a Stroop effect due to the interference generated by the automaticity of note processing in musicians. By trading word reading for note naming, the musical Stroop paradigm provides much 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 in further studies. However, to be a useful tool of investigation, and not only a funny attraction, the musical Stroop effect (1) must fulfill a few additional criteria and (2) must have specific advantages with regard to other, nonstandard Stroop-like procedures for research on automatisms. These two points will be discussed in turn.

Is the Musical Stroop Effect a Manageable Tool of Investigation?

Using the musical Stroop effect requires that the potential study population is large enough. Musicians meeting the usual criteria of expertise for experimental studies on music cognition (in terms of years of practice in music school) are, admittedly, a small fraction of the general population. For instance in France, this fraction is estimated to about 2%. Given that the advantages linked to the musical Stroop effect (e.g., the possibility of a control group) imply to focus on abilities owned by a selective part of the general population, the only relevant question is: Although considerably reduced in percentage, is the remaining sample large enough for any practical purpose? Even a cursory estimation leads to a positive response. For the sake of illustration: A university comprising 20,000 students should include about 400 potential participants, a pool that is incomparably larger than the pool of patients typically available for neuropsychological investigations.

A second prerequisite is that this selective pool of participants does not differ from the general population along dimensions other than the amount of musical training. In particular, musicians could be more sensitive to interference than nonmusicians irrespective of the source of incongruency, leading to artificially inflate the differences between both groups. This specific argument does not hold, however. Existing evidence suggests that overall musicians are *less* sensitive to interference and have better executive control compared to nonmusicians (e.g., Bialystok & DePape, 2009; Travis, Harung, & Lagrosen, 2011). Although the limitations of the quasi-experimental distinction between the natural groups of musicians and nonmusicians make it difficult to strictly rule out the possibility of confounded variables (as in any other Stroop-like paradigms exploiting abilities acquired in real-world settings), our results support the assumption of interference due to extensive musical training.

A third prerequisite for an easy exploitation of the musical Stroop effect is that this effect is sufficiently robust and easily reproducible. The failure of Zakay and Glicksohn (1985) to observe a reliable effect in conditions conceptually similar to those of Experiment 2 (see Introduction) could suggest that the musical Stroop effect is an elusive phenomenon. We believe that this inference would be unwarranted, however. A part of the explanation for Zakay and Glicksohn's failure to get significant effects may stem from the fact that nonsignificance is inferred from a Scheffé's procedure, which is known to be highly conservative. What would have been the results of planned comparisons remains a matter of speculation. Irrespective of the method of analysis, it remains that Zakay and Glicksohn's study lacked power, with a single response time per condition (i.e., for a homogeneous sequence of ten notes) for each participant. In addition, it is also worth noting that Zakay and Glicksohn's study did not meet the current methodological standards in Stroop research. Among several other points,[3] all the notes were displayed as a continuous sequence on a sheet of paper. This latter feature is typical of early Stroop studies, in which RTs were measured with a manual chronometer for a whole sequence of stimuli belonging to the same condition. All recent experimental studies (including the present one) used a computerized item-by-item mode of presentation with mixed conditions, a procedure that minimizes the use of explicit strategies (MacLeod & MacDonald, 2000) and allows, among other advantages, to remove RTs for errors from RT analyses.

Meeting current methodological standards, we observed a reliable musical Stroop effect in two independent experiments using different procedures. Although this effect was numerically small (within a 8–10 ms range), Cohen's *d*s indicated moderate (Experiment 2) to large (Experiment 1) effect sizes according to the conventional benchmarks. In addition, the numerical size of the effect may have been lowered by several factors. One of them is the proportion of congruent trials. In the standard color-word version, it has been shown that interference increases with the proportion of congruent trials (Lowe & Mitterer, 1982). In our experiments, the proportion of congruent trials was small (20%) due to the introduction of various control conditions (out-of-context notes and non-note words). Our results suggest that introducing these additional conditions in future investigations exploiting the musical Stroop effect does not look as mandatory, and we conjecture that a procedure involving only the two main experimental conditions, with half of the trials being incongruent and the other half congruent, would result in larger effects.

Arguably, the effect size would also be increased if a smaller sample of note positions was involved. In the present procedure, each note could be located on two positions on the staff in the congruent condition, hence exploring the knowledge of two full octaves. Using, say, only a few locations surrounding the treble key line would certainly tap the most automatic associations between note names and note positions.

Comparing the Musical Stroop Effect With Other Stroop-Like Paradigms

In the introductory section, we have presented the advantages of the musical Stroop task for studying automatisms by contrast with the standard color-word or picture-word Stroop tasks. Although these standard tasks are prevailing in the literature, there are a number of other variants, which have to be examined.

A first category of tasks follows the so-called "reverse Stroop" paradigm, in which the roles of word reading and color naming (or picture naming) are reversed. The stimuli are unchanged, but subjects are instructed to read the words, and any evidence of interference ought to be attributed to the automaticity of the processes engaged to deal with the irrelevant dimension. Since Stroop (1935), the Reversed Stroop Effect (RSE) has been known as an elusive phenomenon. By and large, an RSE has been obtained with oral responses if the usual conditions of reading were strongly degraded (e.g., Dunbar & MacLeod, 1984), or if participants were overtrained to inhibit the word prior to the experimental session (e.g., Dulaney & Rogers, 1994). Note that a reliable RSE has been obtained when verbal responses are replaced by responses made via word-labeled buttons (e.g., Blais & Besner, 2006, 2007). Irrespective of the size and robustness of the RSE, however, abilities such as color naming and picture naming are certainly not good candidates for studying the formation of automatisms, given the difficulty of controlling what counts as "practice" for these tasks in real-world settings.

Other paradigms do not involve reading at all. For instance in studies involving numerosity, counting is impaired whenever the stimuli being counted are incongruent numerals (e.g., Heine et al., 2010; Shor, 1971). But manipulating practice with number is hardly easier than manipulating reading practice, because everyone acquires both abilities at school within the same age span. Other studies have attempted to circumvent the problem by using arbitrarily structured tasks in which training is entirely performed in the laboratory context (e.g., Regan, 1981). This indeed allows a perfect control over practice, but at the expense of losing what makes the Stroop task so attractive, namely its propensity to assess the consequences of very extensive practice over years.

A last to-be-examined paradigm seems, at first glance, more promising. Several studies have used a standard Stroop paradigm except that words belonged to a second language. Second language learning often occurs later than reading acquisition, involves extensive practice, and arguably, the amount of practice can be controlled as well as the learning of music. We do not intend to deny the interest of relying on the learning of a second language to research on automatisms, as successfully performed in a few earlier studies (e.g., Tzelgov, Henik, & Leiser, 1990). However, it is worth noting that the bulk of the Stroop literature involving a second language is devised to enlighten various issues on bilingualism, rather than to improve our understanding of automatization. Of course, as such, this does not mean that the paradigm would be unsuitable to studying automatization, insofar as the current focus on bilingualism could only reflect researchers' main interest. However, the fact that many issues on bilingualism are currently unsolved is actually damaging for drawing clear conclusions on automatisms. To illustrate: What is assumed to be "automatized" with the practice on a second language? Is it the link between the words of each language (e.g., RED-rojo, with rojo activating RED and RED activating an interfering response)? Is it the direct access to the concepts (with rojo directly activating an interfering response)? A response to

[3] In particular, there was no counterbalancing or randomization (except for task order), which may have introduced noise. The presentation of the stimuli, such as that shown in Zakay and Glicksohn's (1985, p. 417), Figure 1 is also questionable. The locations of the names of the notes were somewhat imprecise, with some overlaps. In addition, the note names were not written inside a note picture as in our materials, but simply written at note locations, which might have downplayed the effect.

these questions depends on theoretical options (see especially the "word association" model vs. the "concept mediation" model of bilingual memory, e.g., Kroll & Stewart, 1994). At a more elementary level, what is the role of learning the grapheme/phoneme mapping? This role is presumably null whenever the languages share most of their alphabets (e.g., English/Spanish), but the response is less obvious when the two languages involve different symbols, as English and Chinese (see, for instance, Chen & Ho, 1986). Presumably, the weight of all these factors changes throughout practice, hence making the study of a putative "automatization" process all the more complex. In any case, it would be a huge oversimplification to consider that using a Stroop paradigm with a second language allows reproducing what occurs initially with one's first language, with the advantage of decoupling age and level of practice.

A Final Comment About Reading

Although we focused throughout this paper on the automatism of note naming, it is worth adding that our study is endowed with interesting implications with regard to the automatism of reading. As mentioned above, earlier studies having explored whether reading could be impaired by an incongruent context (i.e., the reverse Stroop paradigm) have mostly concluded in a negative way. This is generally interpreted as evidence that "the extreme automaticity of word reading is very difficult to overcome" (MacLeod, 1998, p. 207). In this context, our positive results appear as an exception. Reading times were slowed down by incongruent note positions, even though note names were clearly printed, in a font that certainly surpassed in size the fonts used in most books or newspapers, and without pretraining to inhibit word processing. This suggests that the extreme automaticity of word reading does not protect it against interference, provided that the competing process is itself strongly automatic.

The intriguing possibility that automaticity of behavior does not make it immune to interference is strengthened by the results from a recent paper by Akiva-Kabiri and Henik (2012). In this paper, musician participants, among other tasks, 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). The most relevant part of the results for the present concern was that absolute pitch possessors showed a significant congruency effect. This effect demonstrates that pitch identification in absolute pitch possessors (a mainly inherited ability) was impossible to suppress, thereby interfering with word reading.

Interestingly, in the Akiva-Kabiri and Henik's (2012) experiment, absolute pitch possessors were unaffected by the written note name when they were asked to label the auditory tone, as if the congruency effect could act only with a single dimension at a time. Whether such an effect is also obtained with the musical Stroop paradigm, in which the automaticity of note naming is due to extensive practice, remains to be investigated. More generally, the possibility of putting two automatisms against each other should open a new way of investigation for further research.

Acknowledgment

This research was supported by Grant No. ANR-09-JCJC-0129.

References

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

Bialystok, E., & DePape, A. M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 565–574.

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

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

Catling, J. C., Dent, K., Johnston, R. A., & Balding, R. (2010). Age of acquisition, word frequency, and picture-word interference. *The Quarterly Journal of Experimental Psychology, 63*, 1304–1317.

Chen, H.-C., & Ho, C. (1986). Development of Stroop interference in Chinese-English Bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 397–401.

Dulaney, C. L., & Rogers, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 470–484.

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*, 622–639.

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

Heine, A., Tamm, S., De Smedt, B., Schneider, M., Thaler, V., Torbeyns, J., . . . Jacobs, A. (2010). The numerical Stroop effect in primary school children: A comparison of low, normal, and high achievers. *Child Neuropsychology, 16*, 461–477.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 39*, 149–174.

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

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition, 7*, 485–495.

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

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

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

Nealis, P. M. (1973). The Stroop phenomenon: Some critical tests of the response competition hypothesis. *Perceptual and Motor Skills, 37*, 147–153.

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

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

Regan, J. E. (1981). Automaticity and learning: Effects of familiarity on naming letters. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 180–195.

Schulz, T. (1979). Components of the reaction time Stroop-task. *Psychological Research, 40*, 377–395.

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

Shor, R. E. (1971). Symbol processing speed differences and symbol interference effects in a variety of concept domains. *Journal of General Psychology, 93*, 281–288.

Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture-word processing. *Journal of Experimental Psychology: General, 109*, 373–392.

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

Tzelgov, J., Henik, A., & Leiser, D. (1990). Controlling Stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 760–771.

Travis, F., Harung, H. S., & Lagrosen, Y. (2011). Moral development, executive functioning, peak experiences and brain patterns in professional and amateur classical musicians: Interpreted in light of a Unified Theory of Performance. *Consciousness and Cognition, 20*, 1256–1264.

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*, 414–423.

Received June 8, 2012
Revision received October 23, 2012
Accepted December 19, 2012
Published online April 4, 2013

Laurent Grégoire

Université de Bourgogne
LEAD-CNRS UMR5022
Pôle AAFE, Esplanade Erasme
21000 Dijon
France
E-mail laurent.gregoire@u-bourgogne.fr

