

Reading music modifies spatial mapping in pianists

LAUREN STEWART, VINCENT WALSH, and UTA FRITH
University College London, London, England

We used a novel musical Stroop task to demonstrate that musical notation is automatically processed in trained pianists. Numbers were superimposed onto musical notes, and participants played five-note sequences by mapping from numbers to fingers instead of from notes to fingers. Pianists' reaction times were significantly affected by the congruence of the note/number pairing. Nonmusicians were unaffected. In a nonmusical analogue of the task, pianists and nonmusicians showed a qualitative difference on performance of a vertical-to-horizontal stimulus-response mapping task. Pianists were faster when stimuli specifying a leftward response were presented in vertically lower locations and stimuli specifying a rightward response were presented in vertically higher locations. Nonmusicians showed the reverse pattern. No group differences were found on a task that required horizontal-to-horizontal mappings. We suggest that, as a result of learning to read and play keyboard music, pianists acquire vertical-to-horizontal visuomotor mappings that generalize outside the musical context.

Introspecting on the process of sight-reading, pianist Boris Goldovsky says:

I have come to the conclusion that it [sight-reading] has something to do with the speed at which the visual image is converted into a muscular act; and with some people the transformation is so rapid that a great deal of what happens escapes awareness—it just happens.

(from Wolf, 1976)

Just as reading written text appears to be automatic for literate individuals, performing music from a written score appears to be automatic for those who have a reasonable degree of musical literacy, enabling them to play complex pieces of music at first sight. However, to date, the presumed automaticity of music reading has not been empirically demonstrated.

Automaticity of word reading, in contrast, has been shown repeatedly by the classic Stroop task, in which printed names of colors interfere with naming of the ink color (e.g., the word RED written in blue ink), while ink color does not interfere with color word reading (MacLeod, 1991; Stroop, 1935). Conversely, facilitation occurs when both dimensions of the stimulus specify the same response (e.g., the word RED written in red ink) (Glaser & Glaser, 1982; MacLeod & Dunbar, 1988). The two main theoretical accounts of the Stroop effect are the *relative speed* account and the *automaticity* account. Briefly, the

relative speed account (Morton & Chambers, 1973; Posner & Snyder, 1975) assumes that interference occurs at the response stage and that the direction of interference is determined by the relative speeds of each of the potential responses (one from the ink color, one from the color word). Since word reading is faster than color naming (Cattell, 1886), it is argued that word reading will always interfere with color naming, but not vice versa. The automaticity account focuses on the extent to which each process demands attention. Naming the ink color is assumed to draw more heavily on attentional processes than word reading does. Reading the word is seen to be obligatory, whereas naming the ink color is not. The speed of processing and automaticity accounts are not mutually exclusive; speed of processing and automaticity of a process will often, but not always, correlate, and both will be determined by experience. One model (Cohen, McClelland, & Dunbar, 1990), however, incorporates aspects of both the relative speed and the automaticity accounts. Its emphasis is on the strength, and not the speed, of processing. Each dimension of the stimulus (the color word and the ink color) provides input to a common response layer, and task demand units specify the dimension on which the response should be based (e.g., “respond according to ink color” or “respond according to color word”). When the information from both dimensions is incongruent (e.g., the word RED written in blue ink), response selection is delayed. When information from both dimensions is congruent (e.g., the word RED written in red ink), response selection is facilitated (but see MacLeod & MacDonald, 2000, for an alternative explanation of facilitation effects).

In order to ask whether the presence of irrelevant musical notation has a measurable effect on the speed at which pianists execute a sequence of number to finger mappings, we used a novel musical Stroop task that was based on the

This work was supported by Medical Research Council Grant G9617036; U.F. V. W. is supported by a Royal Society University Research Fellowship. We thank John Morton, Sam Gilbert, Lucy Mills, Steven Dakin, William Thompson, and three anonymous reviewers. Correspondence may be addressed to L. Stewart, Institute of Cognitive Neuroscience, Alexandra House, 17 Queen Square, London WC1N 3AR, England (e-mail: l.stewart@ucl.ac.uk).

classic language Stroop task. In our musical Stroop task, numbers, referring to the fingers of the right hand, are superimposed onto written musical notes, and participants execute a sequence of keypresses by mapping from numbers to fingers. Thus the numbers are the analogue of the ink color, the relevant dimension, in the language Stroop task, and the notes, the analogue of the word, the “to-be-ignored” dimension. The hypothesis is that pianists possess two mappings: the mapping between musical notes and fingers, and the mapping between numbers and fingers. The former will be characterized by a greater strength of processing than will the latter, and it will interfere at the response selection stage. Nonmusicians, having only the number to finger mapping, will not show interference from the musical notation. Note that it is possible that the number-to-finger mapping will be stronger for the pianists than for the nonmusicians, since pianists often use numbers to indicate fingering in a piece of music. However, this would suggest that in pianists, the two types of mapping would be closer together in terms of their strength of processing, arguing against the hypothesis that musical notation will interfere.

In Experiment 2, we used a nonmusical analogue of the musical Stroop task to test a hypothesis concerning the nature of the representation of musical notation. Musical notation is a system that maps pitch height to vertically organized spatial locations (specified by a set of horizontal lines called a *staff*), whereby high-frequency pitches are represented higher up on the staff than low-frequency pitches. Time is mapped horizontally, so that a sequence of musical events is mapped from left (earlier events) to right (later events) on the staff. For instruments such as the piano, the vertical position on the score also maps onto the horizontal location of the note on the keyboard, so that notes that are visually higher (and higher in pitch) are produced toward the right of the keyboard and notes that are visually lower (and lower in pitch) are produced toward the left. We hypothesized that the representations used during music reading would be characterized by a set of vertical-to-horizontal stimulus–response mappings. If pianists have developed such spatial mappings, they may be evident on tasks that involve stimulus–response mappings similar to those used in music reading but that lack any surface resemblance to a music reading situation.

EXPERIMENT 1 Musical Stroop Task

The effect of irrelevant musical notation on performance of the task was measured by comparing response times for executing the motor sequence under different conditions of number/note congruence.

Method

Participants. Two participant groups were used: 12 pianists (10 female, average age 26), recruited from the Royal Academy of Music and 14 nonmusicians (10 female, average age 22), recruited from within University College London. On average, the participants in the pianist group had been playing piano for 20 years (range:

13–34 years). The participants in the nonmusician group had no experience of reading or playing music. All participants were healthy and right-handed and gave their informed consent. The participants were naive to our experimental hypotheses.

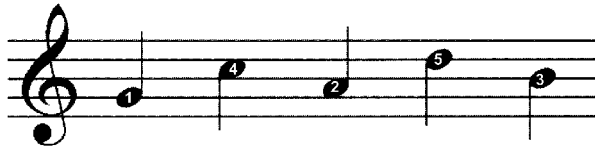
Stimuli. Stimuli were generated using the music notation software package Sibelius (Sibelius Group, <http://www.sibelius.com>). Each stimulus was a notated bar of five crotchets (quarter notes) on which numbers were superimposed. Each note (G, A, B, C, D) and each number (1, 2, 3, 4, 5) appeared once in every stimulus. Five types of stimuli were used (Figure 1). *Baseline* stimuli contained no musical notation and consisted of a row of five white numbers against a background black strip. *Congruent* stimuli were musically congruent and therefore also spatially systematic: G, the lowest note used, would be labeled “1”; A, the next lowest note used, would be labeled “2”; and so on. This mapping is similar to that used by keyboard players reading a score; notes extending from the bottom to the top of the staff map respectively onto digits extending from the left to the right of the hand. *Incongruent (random)* stimuli were musically incongruent and spatially unsystematic: B might be labeled “4”; G might be labeled “5”; C might be labeled “1”; and so on. The relationship between notes and numbers was not consistent across stimuli of this type, so the participants could not learn an arbitrary number/note pairing. The incongruent (random) stimuli were designed to exhibit an approximately equal degree of note/number incongruence across the set. The method of determining note/number incongruence was as follows: If the specified number to finger mapping was, for example, 13542, each pair of consecutive elements of the sequence was considered separately (e.g., 1–3, 3–5, 5–4, 4–2) and the direction and magnitude of each difference was calculated. The first pair, 1–3, shows a positive difference with a magnitude of 2; thus, these numbers would be superimposed onto a pair of notes opposite in their direction (proceed from high to low on the staff) and separated by a magnitude other than 2. The possible note pairs were calculated for each of the number pairs, yielding several possibilities. The choice of note pairs for an entire sequence was made so that each of the notes G to D was used only once in combination. *Incongruent (systematic)* stimuli were musically incongruent but spatially systematic: the number/note relationship was the inverse of that used in the congruent trials: G, the lowest note used, would be labeled “5”; A, the next lowest note used, would be labeled “4”; and so on. Since stimuli were either completely congruent or completely incongruent, we used *catch* trials to deter participants from using a note-reading strategy on congruent trials. These stimuli were part congruent and part incongruent (random): Two of the last three elements would be unpredictably incongruent. A music-reading strategy on these trials would therefore be maladaptive. Twelve motor sequences were used, each of which could be presented as baseline, congruent, incongruent (random), incongruent (systematic), or catch trials. Each trial was presented twice, making a total of 120 trials.

Task. Participants sat with the fingers of the right hand resting over five adjacent keys of a laptop keyboard. A computer keyboard was used rather than a piano keyboard, in order to equate familiarity with the response interface across the two groups. The stimuli were presented and the responses recorded using a program written in MATLAB (The Mathworks, <http://www.mathworks.com>). Each trial consisted of one stimulus presentation. Before each trial, a central fixation point appeared for 1 sec. Following this, the stimulus was presented in the center of the screen for 3 sec or 4 sec for the pianist and nonmusician groups, respectively (timings were based on a pilot study), after which time the fixation point reappeared for another second before the second stimulus appeared. The participants were instructed to ignore the musical notation and use only the numbers to perform the task. The number “1” inside a note indicated that the participant should press the key beneath the thumb, the number “2” indicated a keypress beneath the index finger, and so on. The participants were instructed to read from left to right, mapping the number information presented inside each note to the correct key as

Baseline



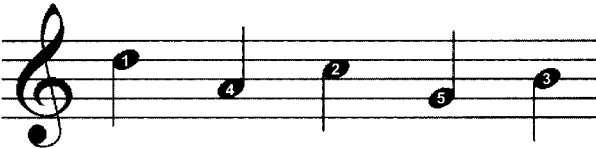
Congruent



Incongruent (random)



Incongruent (systematic)



Catch

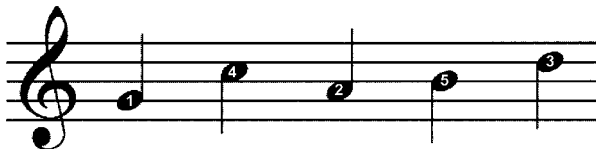


Figure 1. Each motor sequence (12 in total) could appear with five different number/note arrangements. Here each trial type is shown for one motor sequence (1, 4, 2, 5, 3): baseline, congruent, incongruent (random), incongruent (systematic), and catch. Motor sequences and trial types were pseudorandomly interspersed.

quickly and as accurately as possible. It was stressed that the hand should remain over the same five keys throughout the experiment. A practice session consisting of five trials was given to familiarize the participants with the task. Twelve blocks, each comprising 10 stimuli, were presented. Trials were pseudorandomly ordered across motor sequence and trial type (baseline/congruent/incongruent(random)/incongruent(systematic)/catch). First and second presentations of each trial were also pseudorandomly ordered. Thus the participants did not know in advance what trial type or motor sequence to expect.

Predictions

The pianists' response times were predicted to be shorter on the congruent trials than on the baseline trials. In the case of baseline trials, the response could only be arrived at by using the number-to-finger mapping. In the case of congruent trials, both the number-to-finger mapping and the note-to-finger mapping specified the same response. Although instructed to ignore musical notation, if pianists automatically read musical notation, they would derive facilitation from the congruent notation at the response stage.

The pianists' response times were predicted to be longer on both types of incongruent trials than on the baseline trials. The number-to-finger mapping and the note-to-finger mapping used in the incongruent trials specified different responses. Again, although the instruction was to ignore musical notation, if music reading is automatic, these two conflicting stimulus-response mappings would result in interference at the response stage.

Nonmusicians were predicted to be unaffected by the trial type (baseline/congruent/incongruent/catch). Nonmusicians have no existing pathway from notes to fingers; thus, regardless of the congruence of the number-to-finger mapping and note-to-finger mapping, only the number-to-finger mapping would be used in selection of a response.

The pianists' response times were predicted to be longer on the incongruent (random) trials than on the incongruent (systematic) trials. The incongruent (systematic) trials were characterized by a rule-based relationship between the number and its location on the staff; the incongruent (random) trials were not. The spatial systematicity governing where a number would appear could therefore be used, in addition to the number information, for selection of the appropriate response. The congruent trials were also characterized by a spatial systematicity that governed where the numbers appeared. However, the kind of spatial systematicity contained in the congruent trials, in contrast to that in the incongruent (systematic) trials, corresponded to that used in musical notation. Since nonmusicians have no existing note-to-finger mapping, we predicted that the effect of spatial systematicity in the irrelevant dimension would be negligible.

Statistical Analysis

Errors. Because, in the cumulative analysis (see below), each five-element sequence was treated as a single trial, a trial was discarded if a participant made one or more errors out of a possible five.

Response time. Keypress identities and onsets were collected and used for two types of statistical analysis; *cumulative*, in which the time taken to execute each entire sequence was calculated, and *itemized*, in which the response times for each of the responses (first to fifth) in a sequence were used.

Cumulative analysis. Response time data was processed for each participant separately. Trials on which the participant produced an incorrect sequence were discarded. Since each stimulus was presented twice, the cumulative response time for each trial was averaged across these two presentations. In cases where one of the two trials was discarded because of production of an incorrect sequence, the cumulative response time was taken from the single remaining trial. The cumulative response times were collated into motorically matched sets, where a set comprised each trial type of 1 of the 12 motor sequences. If any set of data was incomplete—for instance, because a participant had produced an error on both presentations of a particular trial type for that sequence—the entire set of data was removed. Although this resulted in loss of data, it was necessary to ensure that any eventual differ-

ences in cumulative response time as a function of trial type (baseline/congruent/incongruent(random)/incongruent(systematic)/catch) would not merely reflect the frequency distribution of certain motor sequences. If, after this process, only 50% or fewer sets of data remained, the participant was eliminated from further analysis. For the remaining sets, cumulative response times were averaged across each trial type. To exclude overall differences in response time as a result of pianists' simply being faster to make the keypresses, cumulative response times for each participant were expressed as a proportion of that participant's cumulative response time for baseline trials.

Itemized analysis. The data corresponding to all the first keypresses of every sequence were considered separately from the data corresponding to all the second keypresses of every sequence, and so on. In each case, the data were subjected to the same procedures as those described for the cumulative analysis: For each participant, response times were averaged across the two trials and collated into motorically matched sets according to trial type; incomplete sets of data were discarded and response times were averaged across each trial type for each participant separately. To exclude overall differences in response time, response times for each keypress and for each participant were expressed as a proportion of the cumulative response time for that participant's baseline trials.

Results

Errors. Descriptive data are shown for each group in Table 1. Errors shown are out of a possible 24. There was no significant effect of trial type on the number of errors produced for either group. Participants were no more error prone on catch trials than on congruent trials, from which we can infer that they were using a number-reading strategy for both of these trial types.

Cumulative response time. Two participants (1 from each group) were excluded because the number and pattern of their errors resulted in the removal of more than 50% of the data sets. Descriptive data and corresponding graphs for the remaining subjects are shown in Table 1 and Figure 2. A repeated measures analysis of variance (ANOVA) of mixed design with trial type (4 levels: congruent, incongruent [random], incongruent [systematic], catch, all as a proportion of baseline) as a within-subjects factor and group (2 levels: pianist, nonmusician) as the between-subjects factor revealed a significant interaction [$F(2.3, 51.6) = 23.96, p < .001$]. Two separate repeated measures ANOVAs, one for each group, with trial type as a within-subjects factor (4 levels) revealed a significant effect of trial type for pianists only [$F(3, 30) = 39.84, p < .001$].

It was necessary to ascertain that any potential systematic bias in the removal of data sets from each group could not account for the group \times trial type interaction described. Although the stimuli were designed so that the degree of congruence/incongruence was equated across all sequences, it was possible that a group \times trial type interaction might have resulted if, after removal of the data

sets containing erroneous responses, the data sets remaining for the pianists were more susceptible to the effects of trial type than those remaining for the nonmusicians. Thus we tested whether the set of motor sequences represented in the pianists' data was more affected by trial type in comparison with the motor sequences represented in the nonmusicians. Because, on occasion, errors were made on both presentations of a given trial, the data set for some participants was incomplete. If a participant was missing data pertaining to more than 3/12 sets of stimuli, the participant was excluded from that specific analysis (0 pianists, 3 nonmusicians). For each of the remaining participants, missing data were replaced with the average value for all the trials of that type. A mixed-design repeated measures ANOVA with sequence (12 levels: each motor sequence) and congruency [4 levels: congruent, incongruent (random), incongruent (systematic), catch, all as a proportion of cumulative baseline response] as within-subjects factors and group (2 levels: pianist, nonmusician) as the between-subjects factor revealed no significant interaction. Neither of the within-subjects ANOVAs performed separately for pianists and nonmusicians revealed an interaction of sequence \times trial type. The original trial type \times group analysis was repeated, excluding the 3 participants who had been removed from the sequence \times trial type \times group interaction because of missing data. The trial type \times group interaction was replicated [$F(3, 36) = 8.02, p < .001$], precluding the possibility that it had resulted from the removal of different data sets from each group.

To explore the interaction further, the pianists' data were subjected to five planned paired samples *t* tests. Significant differences were found between baseline and congruent trials [$t(10) = 2.06, p = .03$], baseline and incongruent random trials [$t(10) = -9.65, p < .001$], baseline and incongruent (systematic) trials [$t(10) = -7.49, p < .001$], incongruent (random) and incongruent (systematic) trials [$t(10) = 2.22, p = 0.02$], and congruent and catch trials [$t(10) = -5.11, p < .001$].

Itemized: Main effect of sequential position (Figure 3, Table 2). In order to test for possible serial order effects, we looked at response times for each position (first

Table 1
Musical Stroop Task: Mean Numbers of Errors (Out of 24) and Cumulative Response Times (in Milliseconds)

Subjects	Errors		Response Time	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Pianists				
Baseline	3.08	2.15	1,559	271
Congruent	2.08	2.68	1,507	225
Incongruent (Random)	3.00	2.83	1,811	335
Incongruent (Systematic)	2.75	2.42	1,760	331
Catch	3.17	2.08	1,644	252
Nonmusicians				
Baseline	4.33	3.28	2,725	387
Congruent	3.50	2.35	2,766	409
Incongruent (Random)	4.75	3.89	2,758	385
Incongruent (Systematic)	4.42	3.23	2,730	392
Catch	4.67	3.37	2,721	353

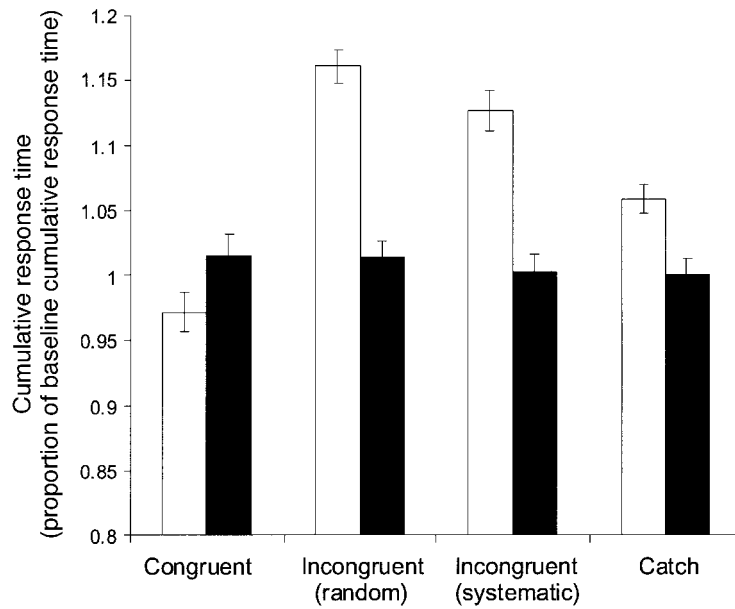


Figure 2. Cumulative response time (time from stimulus onset to final keypress of the sequence) according to trial type [baseline/congruent/incongruent (random)/incongruent(systematic)/catch] for pianists (white) and nonmusicians (black). Data are normalized to the cumulative response time data from baseline trials. Error bars represent ±1 standard error of the mean.

to fifth) in the sequence, regardless of trial type. Response times were averaged across baseline, congruent, incongruent (systematic), incongruent (random), and catch trials. A mixed-design repeated measures ANOVA with sequential position (5 levels: first, second, third, fourth, fifth) as the within-subjects factor and group (2 levels: pianist, nonmusician) as the between-subjects factor revealed an interaction [$F(1.3,26.02) = 27.80, p < .001$], which appeared to arise from the greater drop in response time between the first and subsequent keypresses for pianists than for nonmusicians. Separate repeated measures ANOVAs, with sequential position (5 levels: first, second, third, fourth, fifth) as the only within-subjects factor, revealed a significant effect of sequential position for both groups: pianists [$F(1.23,12.27) = 116.61, p < .001$]; nonmusicians [$F(1.66,16.62) = 136.26, p < .001$]. Although this effect appears to have been mainly due to the relatively long response time taken for the first keypress relative to subsequent keypresses, a significant serial order effect was also seen when the response times were entered from only the second, third, fourth and fifth sequential positions (leaving out the first). This effect was significant in the nonmusicians [$F(3,30) = 23.17, p < .001$] and approached significance in pianists [$F(1.70,16.98) = 3.123, p < .07$].

Itemized: Effect of trial type on sequential position (Figure 4, Table 3). To investigate whether the trial type × group interaction that we previously demonstrated in the cumulative analysis still held for data corresponding to individual responses of the sequence, we performed a

mixed-design repeated measures ANOVA for each sequential position separately, using trial type as the within-subjects factor (5 levels: baseline, congruent, incongruent [random], incongruent [systematic], catch, all as a proportion of cumulative baseline response) and group (2 levels: pianists and nonmusicians) as the between-subjects factor. A significant trial type × group interaction was found for each sequential position: first [$F(4,80) = 12.72, p < .001$], second [$F(4,80) = 3.48, p = .011$], third [$F(2.39,47.8) = 5.34, p = .005$], fourth [$F(4,80) = 2.73, p = .034$], and fifth [$F(4,80) = 2.55, p = .046$]. Separate within-subjects ANOVAs for each group and at each position revealed that nonmusicians showed an effect of trial

Table 2
Musical Stroop Task: Mean Itemized Response Times, Regardless of Trial Type, as a Proportion of Baseline Cumulative Response Times

Subjects	<i>M</i>	<i>SD</i>
Pianists		
Keypress 1	.61	.12
Keypress 2	.12	.04
Keypress 3	.13	.05
Keypress 4	.1	.02
Keypress 5	.1	.02
Nonmusicians		
Keypress 1	.39	.05
Keypress 2	.16	.02
Keypress 3	.18	.02
Keypress 4	.16	.02
Keypress 5	.12	.02

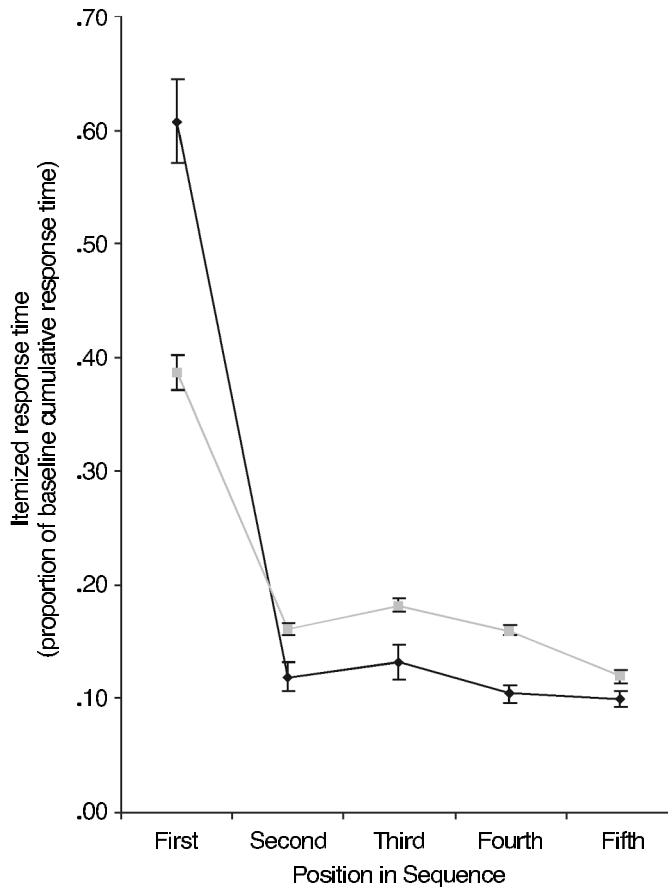


Figure 3. Response times itemized according to sequential position (first, second, third, fourth, fifth), averaged across all trial types. Response time for first sequential position represents the time elapsed between stimulus onset and the first keypress of the sequence. Response time for second sequential position represents the time elapsed between the first and second keypresses, and so on. Pianists are depicted by the dark line; nonmusicians, by the light line. The data are expressed as a proportion of the cumulative response time (stimulus onset to fifth response) for baseline trials. Error bars represent ± 1 standard error of the mean.

type (5 levels) for the first sequential position only. A paired samples *t* test (baseline vs. incongruent [random]) showed that this was carried by a selective increase in response time for incongruent (random) trials [$t(11) = -3.873, p = .003$], suggesting that nonmusicians were sensitive to the lack of spatial systematicity in these trials. Nevertheless, this effect was not seen in the cumulative analysis (see above). The pianists showed a significant effect of trial type (5 levels) at every sequential position: first [$F(4,40) = 28.03, p < .001$], second [$F(1.71,17.13) = 4.57, p = .03$], third [$F(1.46,14.59) = 7.61, p = .009$], fourth [$F(2.19,21.93) = 9.72, p = .001$], and fifth [$F(4,40) = 4.10, p = .007$].

Itemized: Sequential position \times trial type. We assessed whether the effect of trial type differed according to the sequential position of the response. First, we asked whether any potential interaction differed between groups. A mixed-design repeated measures ANOVA with trial type (5 levels: baseline, congruent, incongruent [random],

incongruent [systematic], catch, all as a proportion of cumulative baseline response) and sequential position (5 levels: first, second, third, fourth, fifth) as within-subjects factors and group (2 levels: pianists, nonmusicians) as the between-subjects factor revealed a significant interaction [$F(5.67,113.5) = 4.36, p = .001$]. Separate within-subjects ANOVAs for each group revealed a significant trial type \times sequential position interaction for pianists only [$F(3.33,33.28) = 11.80, p < .001$]. The same analysis carried out in pianists for only the second, third, fourth, and fifth sequential positions (leaving out the first) failed to result in a significant interaction, suggesting that the interaction was carried by a larger effect of trial type on the first response than on subsequent responses.

Discussion

This experiment showed that, for musically literate pianists, the presence of irrelevant musical notation affects the speed at which they can produce a motor sequence,

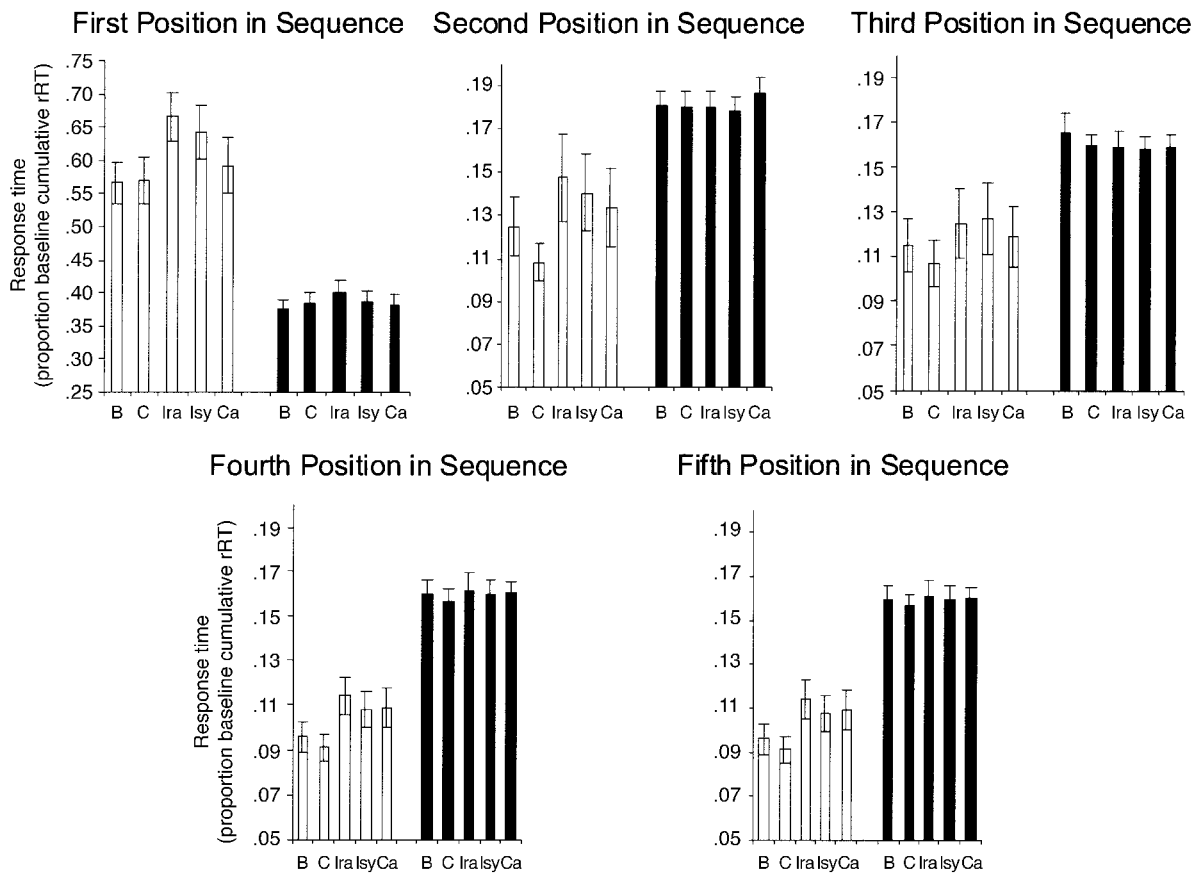


Figure 4. Response times, according to trial type, for each sequential position. Pianists are represented by the white bars; nonmusicians, by the black bars. The data are expressed as a proportion of the cumulative response time (stimulus onset to Key-press 5) for baseline trials. Error bars represent ± 1 standard error of the mean.

using a number-to-finger mapping. In much the same way that decoding from a written word to a spoken output seems to be obligatory (Stroop, 1935), musical notation appears to represent a highly salient and overlearned stimulus for this group of individuals. Nonmusicians, in contrast, are unaffected by the presence of musical notation. Note that this effect is seen despite pianists’ potentially greater experience of mapping numbers to fingers, in comparison with nonmusicians, which would work to minimize any interference effects from the musical notation. This differential effect of musical notation cannot be explained by the removal of a different set of sequences (due to errors) for pianists versus nonmusicians. A comparison of response times for congruent trials and response times for incongruent (systematic) trials revealed a relative interference effect for incongruent trials of the order of 300 msec. Critically, although these two trial types were equated on the degree of inherent spatial systematicity, only the congruent trials made use of the particular spatial systematicity that occurs in musical notation. A separate comparison of congruent and incongruent (systematic) trials, each with baseline, revealed that the congruent – incongruent (systematic) difference comprised both facilitation and interference.

This pattern of data, with interference substantially greater than the facilitation, is typical in Stroop-like tasks (Glaser & Glaser, 1982; MacLeod & MacDonald, 2000). It has been argued that facilitation and interference may not represent opposite sides of the same coin, but rather that facilitation may be an artefact of response error (MacLeod, 1998). Although our data do not pertain to this question directly, the inclusion of catch trials makes it unlikely that the facilitation observed was the result of response error.

When the data were broken down so that we could look at the response times for individual keypresses in the sequence, the overall pattern of response times (averaged across all trial types) showed evidence of response preparation, since time taken to make the first keypress was considerably longer than time taken to make subsequent keypresses. This is consistent with previous work on speech production, which has shown that latency increases linearly with the number of items in the sequence (Monsell & Sternberg, 1981). In addition, there was evidence of a serial order effect for the subsequent keypresses, which was qualitatively similar across the two groups. Notably, response time was increased for the third keypress. Serial order effects have also been reported in

Table 3
Musical Stroop Task: Mean Itemized Response Time,
According to Trial Type, as a Proportion of Baseline
Cumulative Response Time

Position in Sequence	Pianists		Nonmusicians	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
First				
Baseline	.57	.10	.38	.04
Congruent	.57	.11	.39	.05
Incongruent (random)	.67	.12	.40	.06
Incongruent (systematic)	.64	.13	.39	.05
Catch	.59	.13	.38	.05
Second				
Baseline	.11	.04	.17	.03
Congruent	.11	.03	.16	.02
Incongruent (random)	.12	.05	.16	.02
Incongruent (systematic)	.13	.05	.16	.02
Catch	.12	.04	.16	.02
Third				
Baseline	.12	.04	.18	.02
Congruent	.11	.03	.18	.02
Incongruent (random)	.15	.06	.18	.02
Incongruent (systematic)	.14	.06	.18	.02
Catch	.13	.06	.19	.02
Fourth				
Baseline	.09	.02	.16	.02
Congruent	.09	.02	.16	.01
Incongruent (random)	.11	.03	.16	.02
Incongruent (systematic)	.11	.03	.16	.02
Catch	.11	.03	.16	.02
Fifth				
Baseline	.10	.02	.12	.02
Congruent	.10	.02	.12	.02
Incongruent (random)	.11	.02	.12	.02
Incongruent (systematic)	.10	.02	.12	.02
Catch	.10	.03	.12	.02

other transcription tasks such as typing (Sternberg, Monsell, Knoll, & Wright, 1978).

The effect of trial type that was seen in the cumulative analysis was also seen at the level of the individual response times for each position in the sequence: for Sequential Positions 1 to 5, the effect of musical notation on response time was greater for pianists than for nonmusicians. The pianists also showed an interaction of trial type with sequential position, which appeared to arise because the effect of trial type was stronger at Keypress 1 than at the other positions. However, the effect of trial type was not restricted to the first keypress, suggesting that number/note congruency made a contribution, not only during the initial response preparation period but also during the on-line execution of the sequence.

EXPERIMENT 2

Nonmusical Stroop Task

Having demonstrated, in Experiment 1, that musical notation is automatically processed in musically literate pianists, in Experiment 2 we investigated the nature of the representation of musical notation. One of the major differences between music reading and text reading is that music reading almost always involves a response compo-

nent, whereas text reading can, and normally is, performed without explicit mapping to a response system. Thus the automaticity demonstrated in Experiment 1 is likely to act at the interface between perception and action—that is, at the stimulus–response level.

The more compatible a stimulus–response pairing, based on the degree of stimulus–response overlap on some physical or representational dimension, the greater the response time difference that will exist when this stimulus–response pairing is violated (Kornblum, Hasbroucq & Osman, 1990). For instance, it is easy to map a stimulus appearing on the right onto a rightward response (e.g., the right hand) and likewise a stimulus appearing on the left to a leftward response, but there is a reaction time cost associated with mapping a stimulus appearing on the right to a leftward response and vice versa. In contrast, stimuli that appear at different vertical locations (e.g., high/low) do not naturally map onto a particular response in the horizontal meridian. For instance, there does not appear to be a natural correspondence between high and right, low and left, or vice versa. However, the high-right, low-left mapping is precisely that used by pianists when reading music, since musical notes vary in their vertical location on the staff and systematically map onto a horizontal response system (the keyboard) in this way.

In this experiment, we tested the hypothesis that music reading, for keyboard playing, can be characterized by a set of vertical-to-horizontal stimulus–response mappings. A vertical-to-horizontal stimulus–response mapping task was used to measure response time benefits and costs to making a response that was spatially compatible or incompatible with the learned stimulus–response mappings used by pianists reading music. A horizontal-to-horizontal stimulus–response mapping task served as a control task.

Method

Participants. Eight pianists (7 female; average age, 25) and 14 nonmusicians (10 female; average age, 22) took part in the experiment. All participants had also taken part in Experiment 1, and all remained naive to our experimental hypotheses.

Stimuli. In this experiment, we used a nonmusical analogue of the musical Stroop task. The stimuli were numbers presented at different spatial locations. A standard graphics package was used to draw the stimuli. Two versions were used: a vertical-to-horizontal stimulus–response task and a horizontal-to-horizontal stimulus–response task (Figures 5 and 6). In both tasks, five numbers were presented consecutively in five different locations. The stimulus types were the following: baseline, congruent, and incongruent (systematic). Incongruent (random) stimuli were not used because Experiment 1 had already confirmed that spatial systematicity (regardless of musical congruence) affects response time. Thus the comparison of interest occurred between trials equated on spatial systematicity, such as the congruent and incongruent (systematic trials). The number/location relationships were based on those in the analogous stimuli of Experiment 1. In the horizontal-to-horizontal task, congruent trials were characterized by a “1” appearing in the leftmost position, a “2” appearing in the position second from the left, and so on. In the vertical version, congruent trials were characterized by a “1” appearing in the bottommost position, a “2” appearing in the position second from the bottom, and so on. The stimulus–response mapping used for congruent trials in the vertical-

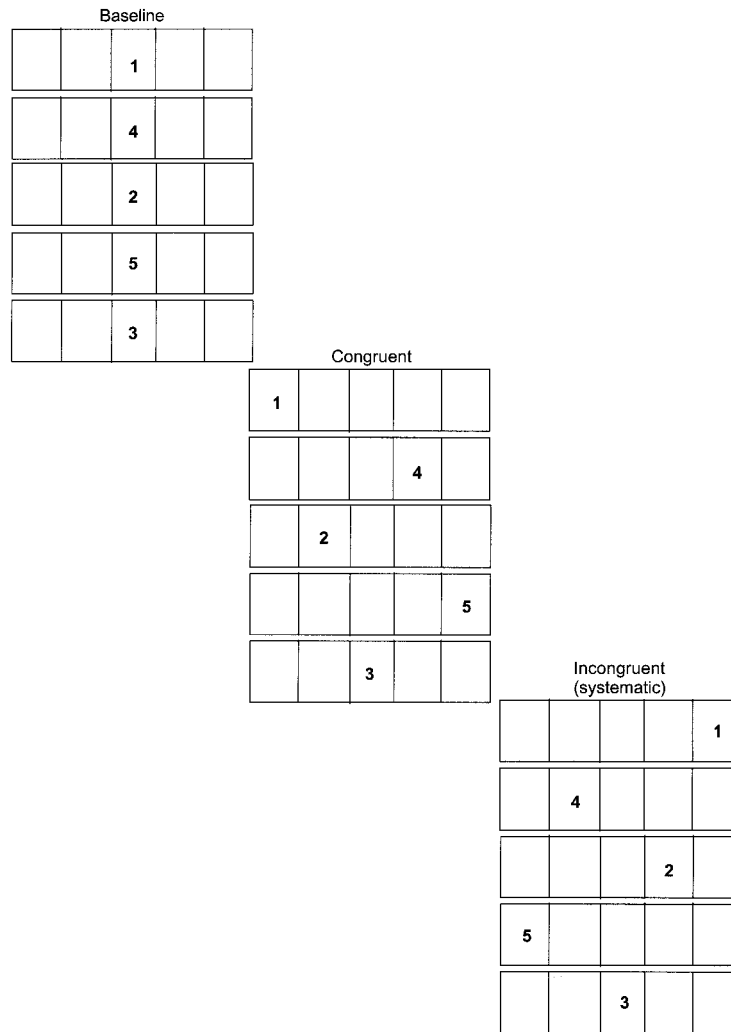


Figure 5. Horizontal-to-horizontal task: Each trial consisted of the sequential presentation of five numbers, within a horizontal array of boxes. Location was to be ignored, and participants performed a simple number-to-finger mapping task. Each keypress triggered the onset of the subsequent number. Here each trial type is shown for one motor sequence (1, 4, 2, 5, 3): baseline, congruent, and incongruent (systematic). Motor sequences and trial types were pseudorandomly interspersed.

to-horizontal version of the task paralleled the mapping used in the reading of keyboard music, in which notes ascending from the bottom to the top of the staff map onto fingers extending from the left to the right of the hand. The incongruent (systematic) stimuli exhibited the inverse relationship to that used in the congruent trials. Whereas the stimuli in the musical Stroop task consisted of five notes presented simultaneously (requiring the participant to proceed serially from left to right), the trials in Experiment 2 consisted of five nonmusical stimuli presented one by one. A keypress made in response to the first stimulus triggered the appearance of the next. This constraint prevented any strategic differences' occurring between the two groups (perceptual or oculomotor) that might account for a performance difference.

Task. In both tasks, the participants placed the right hand over five adjacent keys of a computer laptop keyboard, as in the musical

Stroop task. They were told that a number would appear in one of five locations, and they were instructed to map the number presented onto the appropriate finger and make the required keypress. It was stressed that participants should maintain their hand position above the designated keys throughout the experiment, ignoring any information provided by the spatial location at which the number appeared. The stimuli were presented and the results recorded using a program written in MATLAB. The stimuli appeared one at a time on the computer monitor. As soon as a response was made, that stimulus disappeared from the screen and the next stimulus appeared. After five responses had been made, by each finger once, there was a 1-sec pause before the first stimulus of the next trial appeared. Both the vertical-to-horizontal and horizontal-to-horizontal versions were run, counterbalanced in order across all participants to control for transfer effects between experiments.

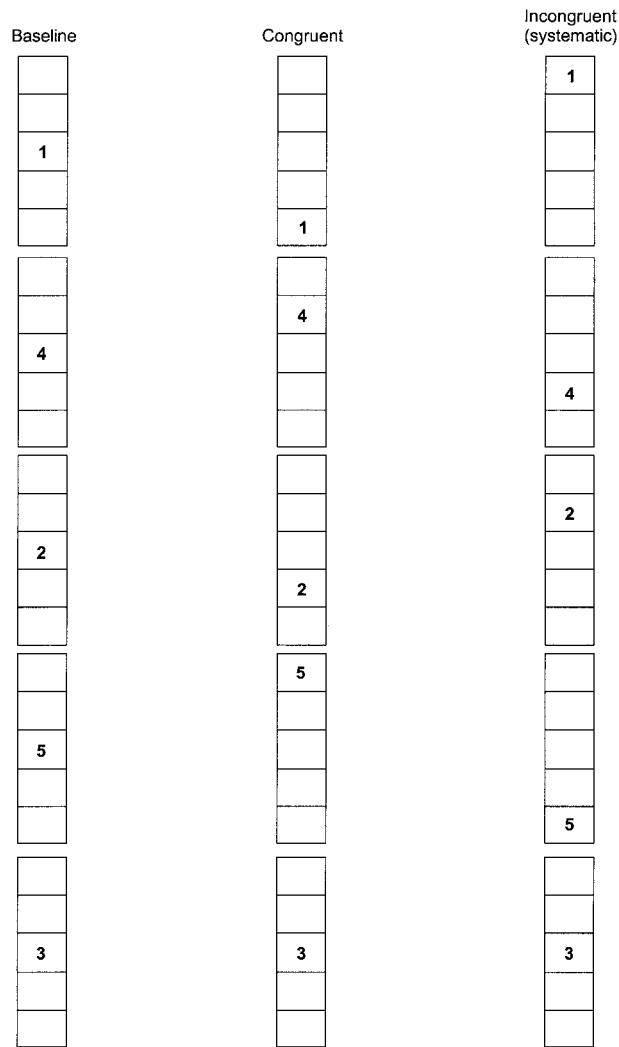


Figure 6. Vertical-to-horizontal task: Each trial consisted of the sequential presentation of five numbers, within a vertical array of boxes. The location at which the number was presented was to be ignored, and participants performed a simple number-to-finger mapping task. Each keypress triggered the onset of the subsequent number. Here each trial type is shown for one motor sequence (1, 4, 2, 5, 3): baseline, congruent, incongruent (systematic).

Predictions

Horizontal-to-horizontal task. It was predicted that both pianists and nonmusicians would have shorter response times on congruent trials and longer response times on incongruent trials than on baseline trials because of the spatial overlap between the location of the stimulus and the required response.

Vertical-to-horizontal task. It was predicted that pianists would have shorter response times on congruent trials and longer response times on incongruent trials than on baseline trials because of their learned association between the vertical position of a stimulus and its horizontal response.

It was predicted that nonmusicians would have similar response times on congruent trials and incongruent trials because, unlike pianists, they would have had no reason to acquire an association between vertical stimuli and horizontal responses.

Statistical Analysis

Response time data were processed for each participant separately. Each set of five consecutive stimuli was treated as a single trial, and each trial was analyzed for cumulative response time. An itemized analysis was not performed. The data were treated in the same way as for Experiment 1: Trials on which the participant produced an

Table 4
Nonmusical Stroop Task: Mean Numbers of Errors (out of 24)
and Mean Cumulative Response Times (in Milliseconds)

Subjects	Errors		Response Time	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Pianists				
Hz to Hz, Congruent	2.00	1.93	2,654	364
Hz to Hz, Incongruent	2.88	3.04	2,792	408
Vt to Hz, Congruent	3.25	2.38	2,785	382
Vt to Hz, Incongruent	4.38	3.20	2,897	419
Nonmusicians				
Hz to Hz, Congruent	2.64	1.60	3,592	475
Hz to Hz, Incongruent	5.36	3.46	3,887	533
Vt to Hz, Congruent	4.36	2.90	3,970	538
Vt to Hz, Incongruent	3.57	2.10	3,814	496

incorrect sequence were discarded, the cumulative response time for each trial was averaged across both presentations, response times were collated into motorically matched sets comprising one of each trial type, and incomplete data sets were removed. For the remaining sets, cumulative response times were averaged across each trial type for each participant separately. This data processing was carried out separately for each task version (vertical to horizontal and horizontal to horizontal). In order to exclude overall differences in response time, response times for each participant were expressed as a proportion of the cumulative response time for baseline trials.

Results

Errors. Descriptive data are shown for each group in Table 4. Errors shown are out of a possible 24 for each trial type. Wilcoxon signed ranks tests revealed that the pianists showed no significant effect of trial type on error rate for either version of the task. The nonmusicians, however, made significantly more errors on the incongruent (systematic) trials than on the congruent trials in the horizontal-to-horizontal task [$Z(-2.64)$, $p = .008$]; they did not differ in the number of errors produced in the vertical-to-horizontal task.

Response time. Descriptive data and the corresponding graphs are shown for each group in Figure 7 and Table 4. A repeated measures ANOVA of mixed design with task version (2 levels: horizontal to horizontal, vertical to horizontal) and trial type [2 levels: congruent, incongruent (systematic), both as a proportion of baseline] as within-subjects factors and group (2 levels: pianists, nonmusicians) as the between-subjects factor revealed a significant interaction of task version \times trial type \times group [$F(1,20) = 13.30$, $p = .002$]. A repeated measures ANOVA with task version (2 levels) and trial type (2 levels) as within-subjects factors revealed a significant interaction for nonmusicians alone [$F(1,13) = 31.43$, $p < .001$].

A repeated measures ANOVA for pianists alone with task version (2 levels) and trial type (2 levels) as within-subjects factors and order of experiments (2 levels: Experiment 1 first, Experiment 2 first) as a between-subjects

factor did not reveal a significant effect of order. Thus no evidence was found to suggest that the pattern of results could be accounted for by transfer effects from Experiment 1 to Experiment 2.

Discussion

The results of Experiment 2 support the hypothesis that musically literate pianists, through extensive practice in reading and playing keyboard music, acquire a set of vertical-to-horizontal stimulus-response mappings that generalize outside of the musical context. Pianists and nonmusicians both showed the expected interference on the horizontal-to-horizontal task, but pianists and nonmusicians showed the opposite pattern of results on the vertical-to-horizontal task. Pianists, as predicted, were facilitated when numbers specifying leftward responses were presented at vertically lower locations and numbers specifying rightward responses were presented at vertically higher locations, the same stimulus-response mapping that is used when a pianist reads music. Nonmusicians, on the other hand, were facilitated when numbers specifying leftward responses were presented at vertically higher locations and numbers specifying rightward responses were presented at vertically lower locations, opposite to the stimulus-response mapping used when pianists read music. Although the pattern of results seen in the nonmusicians was not predicted, it can be explained if we consider the way in which numbers are normally represented in vertical space. When numbers are listed, "1" will usually appear at the top, and subsequent numbers will occupy vertically lower positions. If individuals associate the number "1" with the thumb, the number "2" with the index finger, and so on, their experience of numerical lists may cause them to be quicker to map from a "1" presented in the highest vertical position as opposed to the lowest position. This is to suggest that the effect of trial type seen in the nonmusicians arose because of a learned mapping between space (the experience of numbers ordered vertically from high to low) and digits. If nonmusicians performed the experiment with the left hand, the prediction would be that their responses would be facilitated when small numbers appear in vertically higher locations and larger numbers appear in vertically lower locations. Spatially, this means that, for a nonmusician using the right hand, a leftward response (e.g., pressing the key beneath the thumb [1]) would be facilitated by the stimulus appearing at a vertically higher location, whereas for a nonmusician using the left hand, a leftward response (e.g., pressing the key beneath the little finger [5]) would be facilitated by the stimulus appearing at a vertically lower location. The trial type effect seen in pianists, on the other hand, we argue, arises because of a learned mapping between visual space (experience of notes that vary vertically) and response space (the layout of the keyboard). Thus when pianists perform with either the right or the left hand, leftward responses will be facilitated when the response-eliciting stimulus appears in a vertically higher

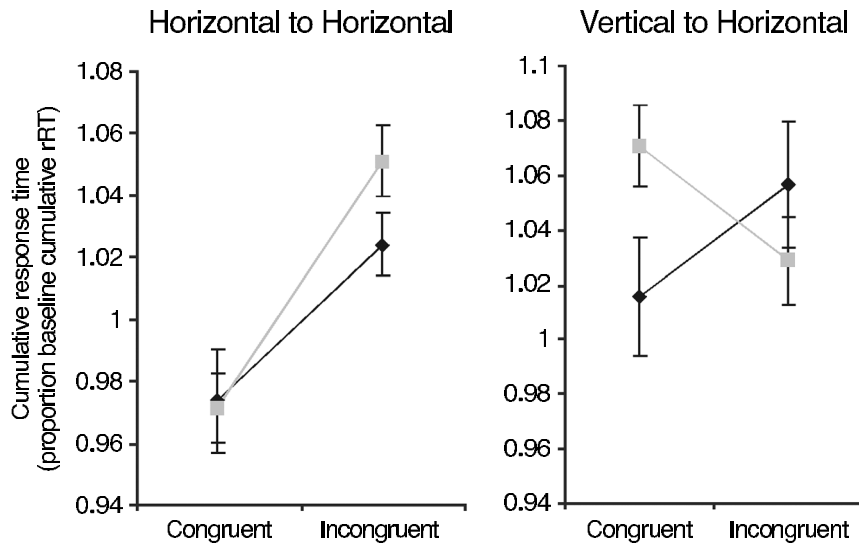


Figure 7. Cumulative response time (time from stimulus onset to fifth response for horizontal-to-horizontal task and vertical-to-horizontal task. Pianists are represented by the dark lines; nonmusicians, by the light lines. Response times are expressed as a proportion of the response for baseline trials. Error bars represent ± 1 standard error of the mean.

location and rightward responses will be facilitated when the response-eliciting stimulus appears in a vertically lower location.

The particular vertical-to-horizontal stimulus–response mappings are likely to be specific to pianists. Other instrumentalists who produce notes by covering certain holes and depressing other keys (woodwind) or by combining fingering with contraction of lip and diaphragm muscles (brass), having had no reason to acquire vertical-to-horizontal stimulus–response mappings, would be unlikely to show equivalent interference on the vertical-to-horizontal Stroop task. However, most conservatory musicians who are not first-study pianists nevertheless possess good keyboard skills, making this a difficult prediction to test.

GENERAL DISCUSSION

Experiment 1 showed that, for musically literate pianists, the presence of irrelevant musical notation can affect the speed at which numbers are converted into a sequence of keypresses. Experiment 2 showed that pianists possess a set of stimulus–response mappings that correspond to the stimulus–response mappings required for one to read and play music. Nonmusicians showed evidence of a different set of stimulus–response mapping, which we attribute to a possible learned association between the vertical organization of numbers in lists and specific digits. The pianists' results from Experiment 2 suggest that musical notation for keyboard performance is represented as a set of vertical-to-horizontal stimulus–response mappings that persist even outside of a musical context. Vio-

lation of these learned mappings is therefore the likely source of interference in the musical Stroop task seen in Experiment 1.

The musical Stroop effect found in Experiment 1 suggests that music reading, like text reading, is automatic. However, as Cohen's model emphasizes (Cohen et al., 1990), automaticity is not an all-or-none phenomenon and is better thought of as a continuum. Automaticity, and thus the interference effects observed in Stroop-like tasks, is subject to the effects of practice. The data on which Cohen's model is based show that 20 h of training on a simple shape-naming task can reverse the direction of interference in a Stroop-like situation (MacLeod & Dunbar, 1988). The pianists in our study had received extensive practice on music reading, averaging 20 years. However, interference effects on the musical Stroop task have also been demonstrated in a group of novice pianists who had been playing the keyboard for only 3 months (Stewart et al., 2003). Whether the automaticity of music reading seen in these novices is as robust as that seen in the highly trained pianists used in our study is a question that remains to be addressed. In addition, longitudinal studies will be required so that one can ascertain the precise point at which interference effects first develop and at what point they reach a maximum level.

The results of Experiment 2 suggest that the acquisition of keyboard skills modifies spatial mapping. Within the music performance domain, knowledge accrued through practice has been shown to transfer at a motoric and/or conceptual level (Palmer & Meyer, 2000), but the effects of music have also been claimed to transfer outside the

musical domain (Chan, Ho, & Cheung, 1998; Graziano, Peterson, & Shaw, 1999; see Schellenberg, 2001, for a review). Reports that music can affect spatial processing emanate from two different types of study: investigations of the nonmusical consequences of listening to music (i.e., the “Mozart Effect”; Rauscher & Shaw, 1998; Rauscher, Shaw, & Ky, 1995; Rauscher et al., 1997), and of the non-musical consequences of musical instruction (reviewed by Hetland, 2000). Studies of the latter kind have reported effects of musical instruction on a wide variety of spatial tasks: spatial-temporal tasks, tasks requiring spatial memory, spatial recognition, mental rotation, and spatial visualization. However, numerous differences among the studies in terms of type and duration of musical instruction, whether or not a control training program was used, the age of the children, and the spatial tasks themselves make it difficult to draw any firm conclusions regarding precisely what is transferred from musical instruction and what conditions are necessary for this transfer to take place. Two kinds of theories have been proposed. “Neural connection” theories such as the “trion” theory (Shaw, 2000) suggest that musical and spatial processing centers in the brain are proximal or overlapping in such a way that the development of certain kinds of musical abilities and that of certain kinds of spatial abilities are related. “Near transfer” theories concentrate on the cognitive similarities between music making and spatial processing. It is proposed that visuospatial intelligence is required for understanding musical notation and spatial relations such as those depicted on keyboards (Hetland, 2000). Thus it is argued that training in music will enhance visuospatial abilities of many different kinds. A meta-analysis of 15 studies of the transfer of musical training to spatial processing (Hetland, 2000) revealed that musical training that included the use of standard notation led to bigger improvements in spatial skills than did musical training that did not. However, even musical programs that did not use notation were found to yield a modest effect on spatial skills. All but two of the studies that used notation also used piano training, making it impossible to say whether either of these variables makes an independent contribution.

In the present experiments, we also investigated the effect of musical training on spatial processing. We did not, however, predict general improvements in spatial abilities; instead we predicted a specific kind of spatial transfer, based on the particular spatial relations used in music reading and keyboard performance. Our prediction that pianists who had learned to read music and play keyboard instruments would possess a set of stimulus–response mappings was supported by the results of Experiment 2. Such a finding suggests that representation of musical notation, at least for pianists, is characterized by specific vertical-to-horizontal stimulus–response mappings. The acquisition of skill in music reading and playing thus involves the building of spatial representations that, once in place, extend outside a musical context.

REFERENCES

- CATTELL, J. M. (1886). The time it takes to see and name objects. *Mind*, **2**, 63-85.
- CHAN, A. S., HO, Y. C., & CHEUNG, M. C. (1998). Music training improves verbal memory. *Nature*, **396**, 128.
- COHEN, J. D., MCCLELLAND, J. L., & DUNBAR, K. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, **97**, 332-361.
- GLASER, M. O., & GLASER, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception & Performance*, **8**, 875-894.
- GRAZIANO, A. B., PETERSON, M., & SHAW, G. L. (1999). Enhanced learning of proportional math through music training and spatial-temporal reasoning. *Neurological Research*, **21**, 139-152.
- HETLAND, L. (2000). Learning to make music enhances spatial reasoning. *Journal of Aesthetic Education*, **34**, 179-238.
- KORNBLUM, S., HASBROUCQ, T., & OSMAN, A. (1990). Dimensional overlap: Cognitive basis for stimulus–response compatibility—A model and taxonomy. *Psychological Review*, **97**, 253-270.
- MACLEOD, C. M. (1991). Half a century of research on the Stroop effect. *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., & DUNBAR, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **14**, 126-135.
- MACLEOD, C. M., & MACDONALD, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, **4**, 383-391.
- MONSELL, S., & STERNBERG, S. (1981). *Speech programming: A critical review, a new experimental approach and a model of the timing of rapid utterances: Part I* (Unpublished technical memorandum). Murray Hill, NJ: AT&T Bell Laboratories.
- MORTON, J., & CHAMBERS, S. M. (1973). Selective attention to words and colours. *Quarterly Journal of Experimental Psychology*, **25**, 387-397.
- PALMER, C., & MEYER, R. (2000). Conceptual and motor learning in music performance. *Psychological Science*, **11**, 63-69.
- POSNER, M. I., & SNYDER, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55-85). Hillsdale, NJ: Erlbaum.
- RAUSCHER, F. H., & SHAW, G. L. (1998). Key components of the Mozart effect. *Perceptual & Motor Skills*, **86**, 835-841.
- RAUSCHER, F. H., SHAW, G. L., & KY, K. N. (1995). Listening to Mozart enhances spatial-temporal reasoning: Towards a neurophysiological basis. *Neuroscience Letters*, **185**, 44-47.
- RAUSCHER, F. H., SHAW, G. L., LEVINE, L. J., WRIGHT, E. L., DENNIS, W. R., & NEWCOMB, R. L. (1997). Music training causes long-term enhancement of preschool children's spatial-temporal reasoning. *Neurological Research*, **19**, 2-8.
- SCELLENBERG, E. G. (2001). Music and nonmusical abilities. In R. J. Zatorre & I. Peretz (Eds.), *The biological foundations of music* (Annals of the New York Academy of Sciences, Vol. 930, pp. 355-371). New York: New York Academy of Sciences.
- SHAW, G. (2000). *Keeping Mozart in mind*. San Diego: Academic Press.
- STERNBERG, S., MONSELL, S., KNOLL, R. L., & WRIGHT, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117-152). New York: Academic Press.
- STEWART, L., HENSON, R., KAMPE, K., WALSH, V., TURNER, R., & FRITH, U. (2003). Becoming a pianist: Brain changes associated with learning to read and play music. *NeuroImage*, **20**, 71-83.
- STROOP, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, **18**, 643-662.
- WOLF, T. A. (1976). A cognitive model of sight-reading. *Journal of Psycholinguistic Research*, **5**, 143-171.