Psychological Research (2018) 82:448–458
DOI 10.1007/s00426-016-0836-4

ORIGINAL ARTICLE

Motor sequence learning and the effect of context on transfer from part-to-whole and from whole-to-part

Zipi Rhein[1] · Eli Vakil[1]

Received: 12 May 2016 / Accepted: 22 December 2016 / Published online: 30 January 2017

Abstract The present study attempts to characterize the contextual conditions (i.e., addition versus omission of elements) that enable or prevent transfer of an acquired skill. The effect of learning and transfer from part-to-whole and from whole-to-part was studied with the serial reaction time (SRT) task. In this study, two alternative sequences of the SRT task were utilized, a short (i.e., 'part') sequence consisting of six elements (ADBACD), and a long (i.e., 'whole') one consisting of 12 elements (BDC**ADBACD**ABC) in which the short sequence was embedded. Three groups participated in the study: one was trained with the 'whole' sequence and two with the 'part' sequence (differing in the number of initial training trials performed), for six blocks followed by a random block. Then, for an additional block, each group was divided into two subgroups, one which continued to practice the same sequence, while the other was transferred to the alternate sequence (i.e., 'part-to-whole' and 'whole-to-part'). Results indicated that the group that first practiced the 'whole' and then the 'part' sequence showed full transfer, while the other group showed only partial transfer from the 'part' to 'whole' sequence. The findings of the present study are inconsistent with Thorndike's principle of identical elements, and, instead, indicate that full transfer is enabled in spite of certain contextual changes (i.e., omissions), but only partial transfer is enabled when other changes are applied (i.e., additions).

Keywords Motor sequence learning · SRT · Transfer · Context

Introduction

The relation between the processes of acquisition and transfer of a learned skill is one of the most important issues that need to be investigated (Speelman and Kirsner 2001). However, the relations between these two aspects of skill learning are not well understood. Extensive training is necessary before a task becomes automatized and performed effortlessly (Adi-Japha, Karni, Parnes, Loewenschuss, and Vakil., 2008; Anderson 1982; Fitts 1964; Karni et al. 1998). Learning is characterized as a power function curve, which reflects several distinct phases of the learning process. The early phases are more controlled and the later phase reaches automatization. A number of studies have demonstrated that the learned skill is more transferable at the early stage rather than at the late stage of training (Rand et al. 2000; Vakil et al. 2002), and therefore, the amount of training is one of the critical factors affecting transfer.

According to Thorndike's principle of identical elements (Thorndike, 1922 in Muller 1999), transfer between tasks is determined by the extent to which they share common elements. Transfer is limited to what was learned previously, and therefore, better performance on the new task will occur only to the extent that it incorporates elements that were part of the learned task. For example, one shared sequence of movements in two different motor skills would yield transfer. Furthermore, the identical elements principle

This study was carried out as part of a PhD dissertation by Zipi Rhein at Bar Ilan University, Ramat Gan, Israel.

✉ Eli Vakil
vakile@mail.biu.ac.il

[1] Psychology Department and the Leslie and Susan Gonda (Goldschmied) Multidisciplinary Brain Research Center, Bar-Ilan University, 52900 Ramat-Gan, Israel

would predict that the larger the shared segment, the better the transfer would be from one skill to another.

However, although this principle sounds intuitively appealing, it is not always supported by the empirical results. The literature that compares whole-task training with part-task training does not always support the identical elements principle that would predict identical transfer under both training paradigms (Speelman and Kirsner 2001). Furthermore, several studies have demonstrated that sequencing performance can be impaired even if none of the elements within the sequence have changed. They showed that performance was impaired even when task-irrelevant features were changed (Ruitenberg et al. 2012a, b).

Several hypotheses were proposed for optimal training and transfer. For example, it has been recommended to prefer part-task learning when the task is highly complex (a large number of task components, e.g., a dance routine) and less organized (degree of dependence of the different elements of the task on one another), and vice versa with whole-task training (Naylor and Briggs 1963; Schmidt and Wrisberg 2008). Their recommendation is to use the whole-task paradigm with a serial task that is highly organized, and the part-task for less organized tasks (see Fontana et al. 2009, for a meta-analysis comparing the various hypotheses). Thus, the relationship between learning and transfer is more complex than assumed by the 'identical elements' principle, because transfer depends on the nature and amount of the training and is not solely determined by the amount of overlapping elements.

The findings that identical elements are not simply transferred from one task to another should not be surprising in light of the evidence that the nature of the task components changes as a function of training and as a function of context. Several studies have demonstrated that as a consequence of extensive training, the performance of the sequence's individual elements is transformed into a new unit or "chunk" that encompasses the learned element and is performed as a single unit (Abrahamse et al. 2013; Verwey and Dronkert 1996). This newly formed unit has different characteristics and is assumed to have a different brain representation than the sum of the individual elements (Sosnik et al. 2004). Speelman and Kirsner (2001) found in their study that the amount of change from the learned skill to the new one is crucial (i.e., one differing calculation versus two or more). Rozanov, Keren and Karni (2010) showed that after extensive training of a finger-movement sequence, even a minor change to one element affected performance on the trained elements that remain unchanged. This finding supports the notion that following extensive training and change of context, the nature of the whole sequence is changed. Speelman and Kirsner (2001) reached the same conclusion, noting that "composed skills appear not to function as encapsulated routines that can be executed in any task context. Rather, composed skills, to some extent, appear to be tied to the context in which they are acquired" (p. 251).

A series of studies (Jiménez 2008; Jiménez et al. 2011) demonstrated that while chunking is typically observed in movement-sequence studies, this is not the case for sequencing performance in the serial reaction time (SRT) task. These researchers show evidence suggesting that certain general statistical information about the sequence structure is learned through training, which leads to better anticipation of the responses.

In the literature comparing whole-task training with part-task training (i.e., all parts that comprise the whole), the criterion for transfer is performance of the whole task. These studies did not test the transfer of a part that is embedded in the whole. Furthermore, transfer in the opposite direction, from whole-task to part-task, was not tested either.

Thus, according to Thorndike's principle of identical elements (Thorndike, 1922 in Muller 1999), transfer is expected regardless of the contextual changes, such as omission (i.e., from whole-task to part-task) or addition (i.e., from part-task to whole-task) of elements to the original learned sequence, because in both cases, there is an identical shared sequence. However, in light of previous findings showing that this principle does not apply under all circumstances (Sosnik et al. 2004), context changes in the form of omissions or additions might hamper transfer. It is important to note that in the present study, the term 'context' refers to an addition or omission from the sequence elements, thus changing the sequence structure and, thereby, the actual movement to be performed. In contrast, previous studies (Abrahamse and Verwey 2008; Ruitenberg et al. 2012a, b) used 'context' to refer to changes in the task setting, whereas the actual sequence that participants performed remained identical to the sequence that they learned during the training phase.

The present study attempts to characterize the contextual conditions (i.e., addition versus omission of elements) that enable or prevent transfer. To this end, two transfer methods were applied: from whole-to-part and from part-to-whole. In the part-to-whole methods, only part of the sequence is identical, but there are new added elements. In the whole-to-part method, certain elements from the whole sequence are omitted, and as before, only part of the sequence is identical. Thus, in both methods, there is an identical shared sequence, but the context differs either by omission or addition of elements to the sequence learned. In a previous study by Verwey (2003a), highly practiced short sequence elements were included in longer SRT sequences, and no transfer was observed.

As described above, the unique contribution of the present study is that the symmetry of transfer will be evaluated under the whole-to-part condition compared to 'whole-continued', and under the part-to-whole condition compared to 'part-continued'. Park, Wilde and Shea (2004) studied sequencing performance following whole-learning compared to part-whole learning. The whole-learning group practiced a sequence of 16 elements (AB) on the first, second and third day. Then, this group was tested on part A (8 elements) and part B (8 elements) separately. The part-whole learning group practiced sequence A on the first day and then went through the same procedure as the whole-learning group on the second and third days. Results showed that performance on the whole sequence was identical for both the groups. However, the part-whole practice group performed more quickly upon transfer of sequence B compared to the whole-practice group, though this was not observed for sequence A. These findings may make a significant theoretical contribution by further elucidating the contextual conditions that support or interfere with transfer of learned skills. The study by Park et al. differs from the present study in several significant ways. First, they tested only two transfer conditions, referred to in the present study as 'part-to-whole' and 'whole-continued'. We tested two additional conditions: 'whole-to-part' and 'part-continued'. The second difference is that in their study as described above, the 'part' was the first 8-element sequence and the 'whole' was the additional 8 elements, yielding a 16-element sequence. In our study, the 6-element 'part' was embedded in the middle of the 12-element 'whole' sequence.

The SRT task, first introduced by Nissen and Bullemer (1987), was used to create 'whole-to-part' and 'part-to-whole' sequences. To avoid the distinctions introduced in the literature between early and late parts (Mattoon 1994), and easy and difficult parts (Clawson et al. 2001), we chose to use the middle elements of the sequence as the 'part' sequence. In this study, there were two alternative sequences: a 'whole' one that contained 12 elements and a 'part' sequence that comprised the six middle elements in the 'whole' sequence. Two groups (an additional 'short-part' group was added to control for the amount of training) were trained with one of two sequences of the SRT task—'part' or 'whole'. Then, each group was divided into two subgroups: one continued with the same sequence, and the other was trained with the other sequence, yielding four groups (see Method section for more details). Evidence of improvement upon transition from the first to the second sequence would be considered an indication of transfer. In other words, faster performance despite the change from 'whole-to-part' or 'part-to-whole' would be interpreted as evidence of transfer.

Method

Participants

Eighty-seven undergraduate students (28 males and 59 females) from Bar-Ilan University (Israel) participated in this study. Based on self-reports, the participants had no history of neurologic or psychiatric disorders. The study was approved, as required, by the Institutional Review Board of Bar-Ilan University. Informed consent was obtained from all participants. For the first seven blocks (six sequence blocks + one random), participants were divided randomly into two groups: one that practiced the short sequence ('part' group) ($n = 42$; mean age 22.61, SD 2.80; mean years of education 13.76, SD 1.34) and another that practiced the long sequence ('whole' group) ($n = 45$; mean age 22.58, SD 2.11; mean years of education 13.53, SD 1.06). The groups did not differ significantly either in age, $t(85) = 0.06$, $p = .96$, or educational level, $t(85) = 0.89$, $p = .38$. Following the random block (i.e., the seventh block following the six sequence learning blocks), participants in each group were divided randomly into two subgroups: one that continues training on the same sequence as before (i.e. training with the 'part' sequence for the eighth block—'part-continued' group, $n = 19$; mean age 22.76, SD 2.39; mean years of education 13.89, SD 1.41), and one that trained with the 'whole' sequence for the eighth block, called the 'whole-continued' group, ($n = 20$; mean age 22.38, SD 2.51; mean years of education 13.40, SD 0.75). The other two groups were transferred to training using the alternate sequence: from 'part' to 'whole' sequence—'part-to-whole' group ($n = 23$; mean age 22.48, SD 3.14; mean years of education 13.65, SD 1.30), and from the 'whole' to 'part'—'whole-to-part' group ($n = 25$; mean age 22.59, SD 2.45; mean years of education 13.64, SD 1.25). Thus, in the eighth block, there were four different groups. The groups did not differ significantly either in age, $F(3, 83) = 0.13$, MSE 5.98, $p = .94$, or educational level, $F(3, 83) = 0.56$, MSE 1.43, $p = .69$ (for description of the paradigm see Fig. 1).

In this paradigm, we equated the number of trials (i.e., 108 key presses) per block for the long (i.e., 'whole') and the short (i.e., 'part') sequences. As a result, in terms of the number of sequences repeated, the 'part' group trained twice as much in the first six learning trials as did the 'whole' group. Thus, a follow-up experiment was designed to equate the amount of training of the two groups by halving the number of repetitions of the short sequence. This way, the number of sequences repeated is equated between the groups rather than the number of trials. Therefore, an additional group ('short part-to-whole') of undergraduate students ($n = 18$; mean age 22.53, SD 2.42; mean years of

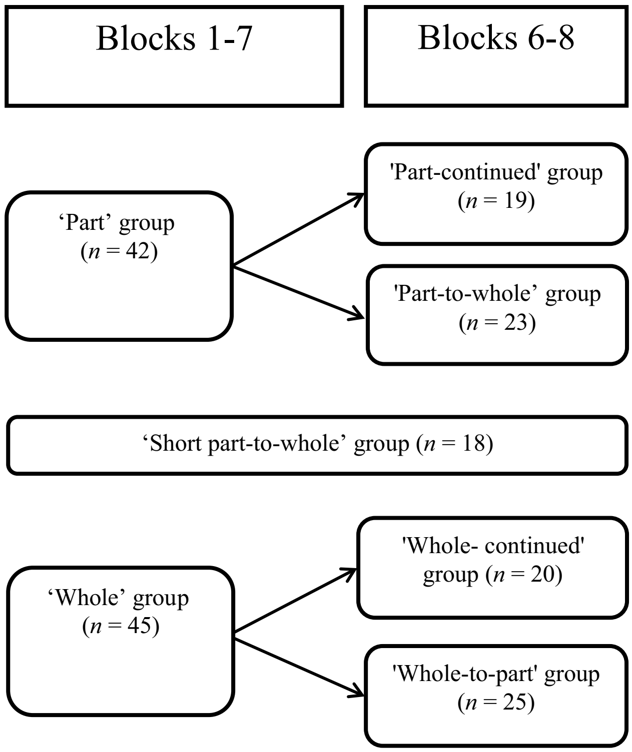

Fig. 1 Description of the paradigm

education 13.17, SD 0.62) from Bar-Ilan University (Israel) was recruited for the follow-up study.

Task and procedure

Participants were tested individually in one session with the SRT task. The participants were seated in front of a PC that ran a SuperLab (Cedrus, Inc.) program application (RB-401) of the SRT task. This program allows measuring the RT to 1000th of a second's accuracy, using a special keyboard with four marked response keys. In a typical trial, a red dot appears in one of four squares (5 × 7 cm) arranged horizontally on the computer screen. Participants were given the following instructions: "A red dot will appear in one of the four elements on the screen. Your task is to press the horizontal numerical key on the keyboard that corresponds with the position of the red dot, as fast as possible, using the index finger of your dominant hand. In other words, identify the location of the red dot from the left-most to the right-most position, and respond by pressing keys A to D, respectively".

In this study, there were two alternative sequences of the SRT task. The 'part' sequence contained six elements—ADBACD, and 18 repetitions of this sequence comprised one block (108 trials). The 'whole' sequence contained 12 elements—BDC**ADBACD**ABC, and nine repetitions of this sequence comprised one block (108 trials). Reed and Johnson (1994) raised the issue of the probabilistic sequence structure. This sequence follows Reed and Johnson's requirement that first-order predictive information is provided (i.e., each location is preceded by the same location only once—AB, AC, AD, BA, BC, etc.).

As can be seen, the 'part' sequence is embedded in the middle of the 'whole' sequence. However, the two different sequence lengths yield unequal probabilistic structures.

Participants practiced the sequences for six consecutive blocks, each of which contained 108 key presses, with a 1-minute break between blocks. The sixth block was followed by a random block that was identical for all participants and all conditions. The random block (#7) also contained 108 trials. The cost in reaction time (RT) for transfer from the sixth block to the random block is considered a purer measure of the sequence learned than improvement in RT from the first to the sixth block (Nissen and Bullemer 1987). This is because the latter reflects general knowledge about the task that is learned (e.g., the S–R mapping of the key response to the corresponding red box on the screen) in addition to learning the specific sequence. As described above, after the first seven blocks, each of the two groups was divided into two subgroups for the eighth block (one continued the same sequence and the other switched to a new sequence). The transfer block (#8) also contained 108 trials.

As soon as a response was recorded, or if the participant did not respond within 5 s, the next target appeared on the screen whether or not the response made was correct. RT was defined as the time from onset of the stimulus to pressing a response key. RT was recorded automatically by the computer for correct responses, and incorrect responses were recorded as errors.

The follow-up group 'short part-to-whole' performed the same test procedure as the 'part-to-whole' group with one difference: they received half the amount of sequence training. In the previous experiment, both the groups had an equal number of learning blocks (i.e., 6), which resulted in practicing the short sequence twice as many times as the long one. The 'short part-to-whole' group was trained only for three blocks versus the six blocks used in the 'whole-to-part' group, but in this experiment, the number of times that the sequence was practiced was identical for both the groups.

Results

All participants responded within the 5-second time limit. As the error rate along the learning blocks was relatively low ($M = 2.64\%$, SD 1.84), analyses were limited to RT measures. In a preliminary analysis of the learning blocks, it was found that the 'whole' group ($M = 2.87\%$, SD 1.83) did not significantly differ from the ‘part’ group in the number of errors made ($M=2.39\%$, SD 1.85), $t(79)=1.18$, $p=.24$.

Sequence learning is expressed in two ways: first, by the rate of reduced RT over the first six blocks of the repeated sequence, and second, by comparing the repeated sequence (i.e., 6th block) with the random sequence (i.e., 7th block). The median RT for each sequence (which was repeated nine times in each block) was obtained, and then, the mean of these nine medians was calculated for each participant per block.

Learning: blocks 1–6

The mean of the median RT in the first six blocks was submitted to a mixed-design ANOVA with group (2) and learning blocks (6) as factors, with the former as a between-subjects and the latter a within-subjects factor. Figure 2 presents the mean of the median RT as a function of groups and learning blocks. There was an overall significant reduction in RT over blocks 1 to 6, $F(5, 85)=18.87$, $p<.001$, $\eta^2 = 0.18$. Overall, the RT of the ‘part’ group was faster than that of the ‘whole’ group, $F(1, 85)=6.84$, $p<.01$, $\eta^2=0.07$. The group-by-block interaction did not reach significance, $F(1, 85)=1.11$, $p=.36$, $\eta^2=0.01$, indicating that the learning rates of the two groups were similar.

The marginally significant overall difference between the ‘part’ group ($M=385.18$, SD 75.33) and the ‘whole’ group ($M=421.60$, SD 97.33) was detected as early as from the first block, $t(85) = -1.94$, $p=.055$. To test whether this difference was due to the sequence learning that had already occurred in the first block, performance on the random (#7) block was compared between the groups. It is assumed that performance on this block reflects a purer RT measure because it does not contain a repeated sequence, and therefore, it eliminates the effect of differential learning of the groups. The comparison indicated that the RT of the ‘part’ group ($M=463.02$, SD 86.34) and the ‘whole’ group ($M=433.87$, SD 83.26) were not significantly different, $t(85) = -1.60$, $p=.11$. Thus, it is safe to conclude that the groups’ RT did not differ a priori.

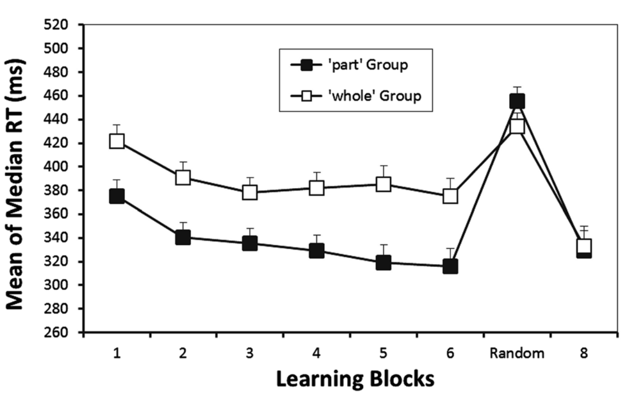

Fig. 2 The mean of the median RT (and SE), of the groups with the ‘part’ and ‘whole’ sequences, in the six blocks of the SRT, the seventh (random) block and the eighth (transfer) block

Learning: block 6 versus block 7 (random)

A comparison of performance in the sixth block and the seventh (random) block was also submitted to a mixed-design ANOVA. As can be seen in Fig. 3, there was an overall increase in RT over blocks 6 and 7, $F(1, 85)=205.86$, $p<.001$, $\eta^2 = 0.71$. Overall, the groups did not differ significantly $F(1, 85)=0.49$, $p=.49$, $\eta^2 = 0.01$; however, the group-by-block interaction reached significance, $F(1, 85)=35.97$, $p<.001$, $\eta^2 = 0.30$. This interaction indicates that the cost of the switch to the random block for the ‘part’ group was greater than the cost for the ‘whole’ group.

Transfer: block 7 (random) versus block 8

To assess the effect of context change on transfer of knowledge, a comparison of performance in the eighth block and the seventh (random) block was also submitted to a 2×4 mixed-design ANOVA. Figure 3 presents the mean of median RT as a function of group (‘part-continued’, ‘part-to-whole’, ‘whole-continued’, and ‘whole-to-part’) and blocks (7th to 8th). This analysis revealed a significant group-by-block interaction, $F(3, 83)=12.19$, $p<.001$, $\eta^2=0.31$.

Follow-up analyses revealed that although all groups showed a significant decrease in RT on the eighth block

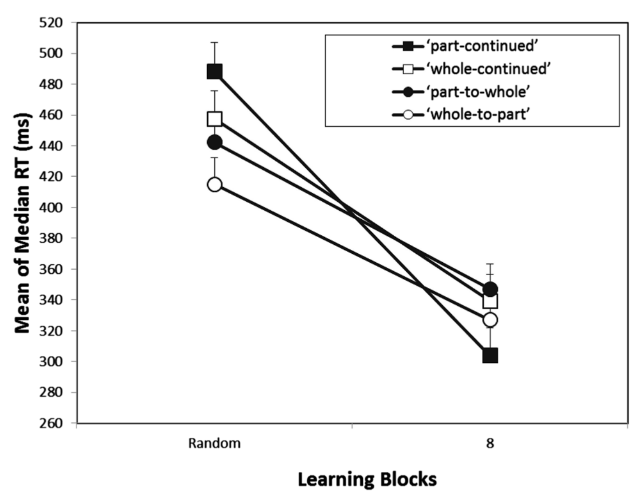

Fig. 3 The mean of the median RT (and SE) of the four groups (‘part-continued’, ‘whole-continued’, ‘part-to-whole’, and ‘whole-to-part’) in the seventh and eighth blocks

compared to the random block, the decrease in RT for the ‘part- continued’ group ($p < .05$) was the steepest.

Transfer: block 6 versus block 8

To assess the effect of context change on transfer of the knowledge acquired during the first six blocks, performance was compared between the sixth and eighth block. A 2×4 mixed-design ANOVA with block (6th versus 8th) and group (‘part-continued’, ‘part-to-whole’, ‘whole-continued’, and ‘whole-to-part’) was conducted with the former being a within-subjects factor and the latter a between-subjects factor. Figure 4 presents the mean of median RT as a function of group and blocks. This analysis revealed a significant group-by-block interaction, $F(3, 83) = 9.64$, $p < .001$, $\eta^2 = 0.26$. To detect the source of interaction, simple analyses were conducted in which the effect of transfer on each group was analyzed separately. The ‘part-to-whole’ group was the only group that showed a significant increase in RT from the sixth to eighth block, $t(22) = 14.89$, $p < .001$. All others showed a significant decrease in RT—‘part-continued’ ($t(18) = 12.95$, $p < .001$); ‘whole-continued’, ($t(19) = 16.91$, $p < .001$); and ‘whole-to-part’ ($t(24) = 19.98$, $p < .001$) (see Fig. 4).

Because the ‘part-to-whole’ group showed that the transition from the ‘part’ sequence to the ‘whole’ sequence involved a cost, follow-up analysis was conducted. The purpose of this analysis was to test the possibility that the cost of transfer was only due to the new elements of the ‘whole’ sequence (in the eighth block) that were not previously learned. It is possible that these new elements led to the increase in RT, while RT continued to decline for the elements learned previously, indicating that there was a transfer of knowledge for these elements. To check this possibility, two new scores were generated for the eighth block—the mean of median RT of the new

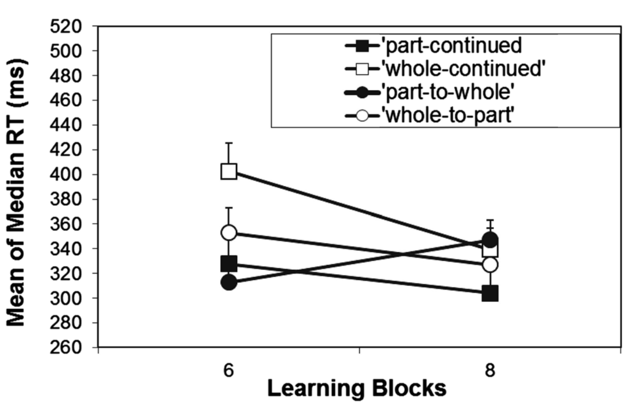

Fig. 4 The mean of the median RT (and *SE*) of the four groups (‘part-continued’, ‘whole-continued’, ‘part-to-whole’, and ‘whole-to-part’) in the sixth and eighth blocks

elements of the ‘whole’ sequence, $M = 348.52$, SD 15.16, and the mean of median RT of the old elements of the ‘whole’ sequence, $M = 347.55$, SD 14.37. As can be seen, these scores are very similar and are not significantly different ($p > .05$). Thus, the cost of transfer is not due to the new elements of the ‘whole’ sequence that were not learned before. Furthermore, this indicates that RT was increased even for old elements, simply because new elements were added to them.

Additional analyses were conducted to better understand the asymmetry of transfer. For this purpose, groups in which the context did not change (‘part-continued’ and ‘whole-continued’) were considered a baseline of sorts, and possible deviations from this baseline were analyzed. Thus, we calculated the differences between performance on block 8 and on block 6 for each of the four groups. The difference scores for the ‘part-continued’ group ($M = 23.65$, SD 63.82) and the ‘part-to-whole’ group ($M = -34.28$, SD 68.76) were significantly different, $t(40) = 2.81$, $p < .01$. The difference score of the ‘whole-continued’ group ($M = 63.13$, SD 58.60) and the ‘whole-to-part’ group ($M = 25.91$, SD 50.41) reached significance as well, $t(43) = 2.81$, $p < .05$.

Thus, the same pattern of transfer was observed for the omission group (i.e., ‘whole-to-part’), that is RT at block 8 was faster than at block 6, compared to the baseline group (i.e., ‘whole- continued’). The significant effect indicates that although the transfer is significant, the extent of significance was less than that under the no-context-change condition. However, for the addition group (i.e., ‘part-to-whole’), the change compared to baseline group (i.e., ‘part-continued’) was reversed. That is, while the ‘part-continued’ group showed continuous improvement (i.e., RT block 6 > block 8), the addition group (i.e., ‘part-to-whole’) showed no transfer (i.e., RT block 6 < block 8).

As explained in the "Method" section, we equated the number of trials per block for the long and the short sequences. As a result, in terms of the number of sequences repeated, the ‘part’ group trained twice as much in the first six learning trials as compared to the ‘whole’ group. Thus, a possible interpretation of our findings is that because the ‘part’ group was trained twice as much in the first six learning trials as compared to the ‘whole’ group, it reached a more advanced stage of practice which prevented transfer. A follow-up experiment was designed to examine this tentative interpretation, by halving the number of repetitions of the short sequence. This way, the number of sequences repeated is equated between the groups rather than the number of trials.

Results of the follow-up study

In an attempt to test the possibility that the double-sequence repetitions led to the different pattern of transfer, the 'short part-to-whole' group replaced the 'part-to-whole' in the analyses of the transfer effect.

As with the original groups, to ensure that the groups' RT did not differ a priori, their performance was compared on the first block and on the random block. The findings were the same as with the original 'part' group. That is, the 'whole' group ($M = 421.60$, SD 97.33) was slower than the 'short part' group ($M = 359.16$, SD 66.95) on the first block, $t(61) = 2.49$, $p < .05$. However, the 'whole' group ($M = 433.87$, SD 83.26) and the 'short part' group ($M = 454.61$, SD 79.81) did not differ on the random block, $t(61) = 0.90$, $p = .37$. Thus, here too, it is safe to conclude that the groups' RT did not differ a priori.

Transfer: block 7 (random) versus block 8

To assess the effect of context change on transfer of knowledge, a comparison of performance between the eighth block and the seventh (random) block was also submitted to a mixed-design ANOVA. Figure 5 presents the mean of median RT as a function of group ('part-continued', 'short part-to-whole', 'whole-continued', and 'whole-to-part') and blocks (7th to 8th). This analysis revealed a significant group-by-block interaction, $F(3, 78) = 12.13$, $p < .001$, $\eta^2 = 0.32$. Follow-up analyses revealed that although all groups showed a significant decrease in RT on the eighth block compared to the random block, the decrease in RT for the 'part-continued' group was greatest ($p < .05$).

Transfer: block 6 versus block 8

To assess the effect of context change on transfer of knowledge acquired during the first six blocks, a comparison of performance between the sixth block and eighth block was carried out. Figure 6 presents the mean of median RT as a function of group ('part-continued', 'short part-to-whole', 'whole-continued', and 'whole-to-part') and blocks (6th to 8th). This analysis revealed a significant group-by-block interaction, $F(3, 78) = 7.45$, $p < .001$, $\eta^2 = 0.22$. To detect the source of interaction, simple analyses were conducted in which the effect of transfer on each group was analyzed separately. It was found that the 'part-to-whole' group was the only group that showed a significant increase in RT from the sixth to eighth block, while all the others showed a significant decrease in RT ($p < .05$) (see Fig. 6).

The results of this follow-up experiment yielded exactly the same results as the original experiment, whether the transfer block (i.e., #8) was measured in reference to the random block (#7) or to the sixth block, which reflects best learning. Thus, regardless of whether training was equated for the number of blocks or the number of sequence repetitions, the results were identical. In summary, the omission (i.e., 'whole-to-part') of contextual elements enables full transfer either when compared to the random block (#7) or to the best-learning block (#6). However, addition (i.e., 'part-to-whole' or 'short part-to-whole') of contextual

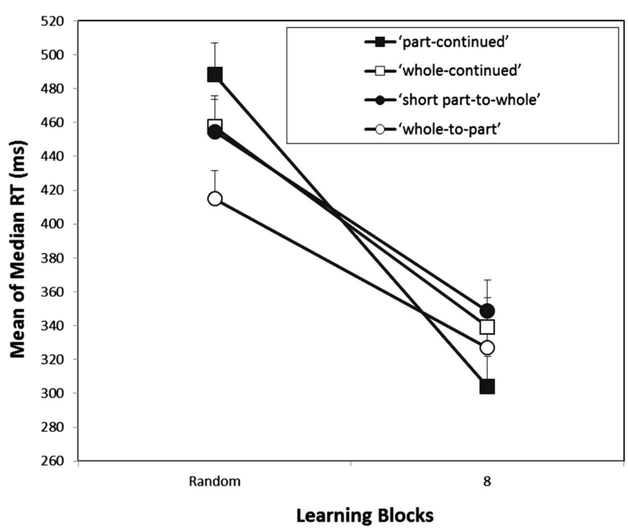

Fig. 5 The mean of the median RT (and *SE*) of the four groups ('part-continued', 'whole-continued', 'short part-to-whole', and 'whole-to-part') in the seventh and eighth blocks (follow-up experiment)

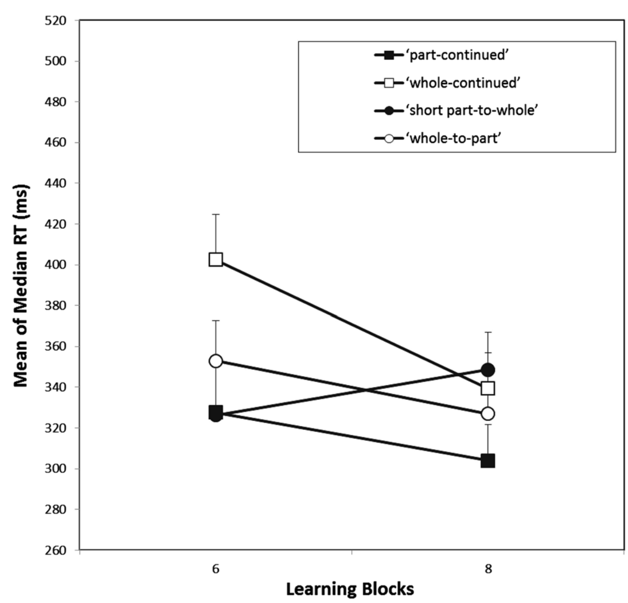

Fig. 6 The mean of the median RT (and *SE*) of the four groups ('part-continued', 'whole-continued', 'short part-to-whole', and 'whole-to-part') in the sixth and eighth blocks (follow-up experiment)

elements enables partial transfer only, because transfer was evident only compared to the random block (#7) but not compared to the best-learning block (#6).

Finally, a direct comparison between the 'part-to-whole' and the 'short part-to-whole' groups in terms of the course of learning through six blocks, random block and whole block transfers were conducted. The results showed the exact same effects for the groups, with no comparisons yielding either group effect or significant interaction with group (all *p*s' >.05).

Discussion

In this study, we attempted to characterize the contextual conditions (i.e., addition versus omission of elements) that enable or prevent transfer. This was achieved by comparing transfer in two contextually changed conditions, that is 'part-to-whole' (or short part-to-whole) versus 'whole-to-part'. In both the methods, there was a shared identical sequence, but the context differed. In the former, transfer is tested when new elements were added, whereas in the latter, certain learned elements were omitted.

The results in the first six blocks indicated that the overall RT of the short (i.e., 'part') sequence group was faster than that of the group trained with the long (i.e., 'whole') sequence. This was also evident by the cost of the transition to the random block. These findings replicate the results of a previous study by Pascual-Leone et al. (1993).

An analysis of the first sequence in the first block was conducted in an attempt to test whether an a priori difference in RT between groups existed. The results showed no significant difference between the groups. This analysis may lack statistical power for interpreting a null effect. Therefore, it is recommended in future research that this part be replicated to avoid concerns of a priori differences, for example, by including a random block at the start. Nevertheless, the fact that the groups' RT did not differ on the random block (#7) indicates that the groups' RT did not differ a priori.

The transfer block was compared with the random block (#8 vs. #7) to test whether there is a transfer of the learning sequence beyond the training effect to more general aspects of learning such as S–R mapping. This analysis showed that all groups showed a decrease in RT, although the 'part-continued' group showed the steepest decrease. Comparison between the sixth and eighth blocks shows faster performance in the transition from 'whole-to-part' (i.e., omissions), while slower performance was observed (see Fig. 4) in the transition from 'part-to-whole' (i.e., additions). Decrease in RT indicates that the modified sequence continues to perform like the original sequence, as expected when transfer takes place. The increase in RT in the 'part-to-whole' group indicates interference and incomplete transfer (i.e., faster than the random block but slower than the best-learning block #6).

As mentioned above, skill acquisition is an extensive process and is composed of several distinct phases that reflect a transition from more controlled to more automatic cognitive processes (Adi-Japha et al. 2008; Anderson 1982; Fitts 1964; Karni et al. 1998). Transfer is better facilitated at the early than at the late stage of training. In monkeys, inter-manual transfer has been shown to be complete early in sequence learning, but incomplete following extended practice (Rand et al. 2000). In humans who learn to solve the Tower of Hanoi puzzle, Vakil et al. (2002) showed that the cost of transfer was significantly greater at advanced, compared with early, stages of learning. Accordingly, while the short sequence training possibly led to automaticity, the long sequence training did not as yet reach that level but rather remained a controlled process, thus enabling transfer.

Korman, Raz, Flash, and Karni (2003) used the finger-to-thumb opposition task to show that transfer of movement-sequence learning is dependent on the quantity of previous practice. They proposed that the early phase of practice may be thought of as perceptual or cognitive repetition priming. However, transfer during the phase of saturation of rapid gains will be less effective, because performance at that stage is routine by nature.

Consistent with these studies, the sequence change from 'whole-to-part' was possible, though at a cost, because training for the long sequence remained at an early phase that enabled transfer to the short sequence. Because training in the short sequence consisted of twice as many repetitions than in the long sequence, participants may have reached a more advanced acquisition stage and, therefore, could not transfer to the longer sequence that followed. In an attempt to test this possibility, a follow-up analysis was conducted with a new group, the 'short part-to-whole' group, which trained half as much as the original 'part-to-whole' group. As reported above, new analyses with the new group yielded results that were identical to the original analyses. This suggests that it is the nature of context change (i.e., omissions or additions) that determines whether transfer occurs, rather than the amount of training.

The findings of the present study are inconsistent with Thorndike's principle of identical elements (Thorndike, 1922 in Muller 1999) which would predict equal transfer from 'whole-to-part' and from 'part-to-whole'. According to this principle, the identical sequence that is shared by both conditions should have led to identical effects in both directions regardless of the contextual changes. However, the results indicate that full transfer is enabled in spite of certain contextual changes (i.e., omissions), and partial transfer is enabled in others (i.e., additions). It should be noted that finding that the 'part-to-whole' (i.e., additions) group was the only group that showed slower RT in block 8 compared to block 6 was not surprising because it was the only group for which new elements were added to the sequence. However, what is surprising and completely contradictory to Thorndike's principle of identical elements is that RT to the old elements in the sequence was the same as to the new elements in the sequence. Furthermore, although the 'whole-continued' and the 'whole-to-part' groups showed faster RT at block 8 compared to block 6, indicating a transfer, the extent of improvement was significantly higher for the former compared to the latter. That is, again, contrary to Thorndike's principle that contextual change (i.e., omission) affects transfer even when the same elements are preserved.

Furthermore, in addition to the finding reported by Speelman and Kirsner (2001), that the amount of change is the critical factor for determining transfer, here, we found that even if the amount of change is identical (i.e., in terms of the number of elements involved), the nature of the context change, addition or omission, was critical as well. These findings show that addition and omission are not symmetrical contextual changes. Adding new elements to a learned skill slows down the execution of old elements. However, eliminating old elements from a learned sequence does not interfere with performance on the new, shorter sequence.

Further analysis of the change from 'part-to-whole' revealed that performance was slowed down not only for the new elements of the 'whole' sequence, but also for old, previously learned elements. As mentioned above, these findings are inconsistent with Thorndike's principle of identical elements. These findings, however, are partially consistent with those of Rozanov et al. (2010), but inconsistent with some of their other findings. In their study, Rozanov et al. used the finger-opposition sequence-learning task with a 5-element sequence. Following intensive training for 3 weeks, they found that even an omission of the penultimate movement resulted in slowing down the initial trained movements. Thus, on the one hand, consistent with the present findings, even a minor change in the sequence affects the execution of the old elements as well. On the other hand, inconsistent with the present findings, the omission of a movement prevented the transfer. One possible explanation for this discrepancy is that in the present study, the 'part' sequence that remains after the omission ('whole-to-part') is identical to that on which participants trained within the 'whole' sequence. However, in Rozanov et al., the penultimate movement of the trained sequence was omitted. As a result, the new sequence of four elements is not only shorter but is, now, also modified because of the transition from the once third element to the fifth element. If this interpretation is correct, omission of the final movement in the Rozanov et al. paradigm would be predicted to enable transfer. The extent of training also differed greatly between the present study (one session) and that conducted by Rozanov et al. (3 weeks). Thus, it is possible that such intensive training would prevent transfer even in a context change of omissions. As described above, the present study is also very different from that of Park et al. (2004), particularly in the change made to the sequence structure between 'part' and 'whole'. Unlike the present study in which the 'part' sequence was embedded in the 'whole' sequence, in the earlier study, the added sequence was placed at the end of the original one.

In order not to interfere with the transfer effect by incorporating declarative knowledge, we chose not to use the generate task in which participants are asked to reproduce the sequence explicitly. However, Pascual-Leone et al. (1993) found that participants showed higher acquisition of declarative knowledge of the sequence when it was shorter. Accordingly, the short sequence probably contained a stronger declarative element, although there was a random block after the short sequence that might obscure declarative knowledge. Previous studies have already demonstrated the importance of declarative knowledge of the sequence for its stability (Averbeck et al. 2006; Brown and Robertson 2007). Thus, it is possible that in the present study, the participants that trained with the shorter sequence had more declarative knowledge of the sequence, which interfered with the transfer.

Consistent with the classic SRT procedure, the learning blocks were followed by a random block (#7) to differentiate between the S–R and sequence learning (Nissen and Bullemer 1987). The transfer block (#8) followed the random block and that itself may constitute interference. Although these blocks enabled us to distinguish between full versus partial transfer, in future research, it is recommended to test transfer directly after the learning phase.

A related interpretation is that the 'part' sequence is within the limits of motor working memory (Verwey 2003), while the 'whole' is not, which facilitated 'chunking' of the sequence. So, regardless of the amount of training, the 'whole' sequence cannot be 'chunked' as a unitary sequence, and, therefore, can be broken down into smaller units with no cost, enabling the transfer from 'whole-to-part'. However, previous studies showed that in the SRT task, what is learned is some general statistical information about the sequence structure rather than chunking (Jiménez 2008; Jiménez et al. 2011). In light of these studies, it can be concluded that the present findings demonstrate that context changes in the form of additions interfere more than omissions with the ability to learn the statistical sequence structure. In conclusion, this study has demonstrated the complex relations between skill acquisition and transfer processes.

One of the limitations of this study is the unequal proportional frequencies of the individual elements in the short versus long sequences. Unlike the short 6-element sequence, the long sequence with 12 elements enables the frequencies to be completely balanced (each element occurring three times), so that it is statistically less redundant. This frequency compound may cause an overestimation of the learning effect with the short sequence (because unconditional stimulus expectancy and sequence learning effects add up). In support of the plausibility of this interpretation is the fact that chunking, which is the alternative strategy to learning the statistical regularities of the sequence, is not an option usually adopted with SRT (Jiménez 2008; Jiménez et al. 2011). Furthermore, the differences between the short and long sequence structure are expressed in effector dependence, and might also have differentially affected explicit knowledge of the sequence. In other words, the short sequence is more likely to yield explicit knowledge than the long sequence. Several studies have showed that explicit knowledge of the task facilitates transfer. Using the SRT task as in this study, Willingham, Nissen, and Bullemer (1989) showed that individuals that demonstrated explicit knowledge of the sequence had faster RT in the transfer task. Ahlum-Heath and Di Vesta (1986) reached a similar conclusion by showing that participants that were required to verbalize their moves while solving the Tower of Hanoi puzzle performed better at the transfer task.

The other limitation is that participants were instructed to use only their index finger to respond. This could have confounded results due to distance and corresponding movement time. For example, if we denote the horizontal keys 1 to 4 from left to right, the distance between key 1 and key 2 is shorter and would likely result in faster transitions, compared to keys 1 and 4.

In summary, regardless of whether the context change involves omission or addition of elements to the learned sequence, some aspects of learning were transferred, as is evident by better performance in both the groups compared to the random block. However, transfer under the additions condition was incomplete (i.e., faster than in the random block but slower than in the best-learning block #6). Thus, full transfer was evident under the contextual change of omission, and partial transfer upon addition. Furthermore, both the groups showed significant cost when compared to transfer of the baseline when context was unchanged. Thus, consistent with Ruitenberg et al. (2012a, b), any change in the sequence would hamper transfer.

Compliance with ethical standards

Funding We declare that this project was not funded. The study was approved, as required, by the Institutional Review Board of Bar-Ilan University.

Informed consent Informed consent was obtained from all participants.

Conflict of interest No conflicts of interest exist.

References

Abrahamse, E. L., Ruitenberg, M. F. L., De Kleine, E., & Verwey, W. B. (2013). Control of automated behavior: insights from the discrete sequence production task. *Frontiers in Human Neuroscience, 7*, 1–16.

Abrahamse, E. L., & Verwey, W. B. (2008). Context dependent learning in the serial RT task. *Psychological Research, 72*, 397–404.

Adi-Japha, E., Karni, A., Parnes, A., Loewenschuss, I., & Vakil. E. (2008). A shift in task routines during the learning of a motor skill: group averaged data may mask critical phases in the individuals' acquisition of skilled performance. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 6*, 1544–1551.

Ahlum-Heath, M. E., & Di Vesta, F. J. (1986). The effect of conscious controlled verbalization cognitive strategy on transfer in problem solving. *Memory and cognition, 14*(3), 281–285.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369–406.

Averbeck, B. B., Sohn, J. W., & Lee, D. (2006). Activity in prefrontal cortex during dynamic selection of action sequences. *Nature Neuroscience, 9*, 276–282.

Brown, R. M., & Robertson, E. M. (2007). Off-line processing: reciprocal interactions between declarative and procedural memories. *Journal of Neuroscience, 27*, 10468–10475.

Clawson, D. M., Healy, A. F., Ericsson, K. A., & Bourne, L. E. Jr. (2001). Retention and transfer of Morse code reception skill by novices: Part-whole training. *Journal of Experimental Psychology: Applied, 7*, 129–142.

Fitts, P. M. (1964). Perceptual - motor skill learning. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press.

Fontana, F. E., Mazzarado, O., Furtado, O., & Gallagher, J. D. (2009). Whole and part practice: A meta-analysis. *Perceptual and Motor Skills, 109*, 517–530.

Jiménez, L. (2008). Taking patterns for chunks: is there any evidence of chunk learning in continuous serial reaction-time tasks? *Psychological Research, 72*, 387–396.

Jiménez, L., Méndez, A., Pasquali, A., Abrahamse, E. L., & Verwey, W. B. (2011). Chunking by colors: Assessing discrete learning in a continuous serial reaction-time task. *Acta Psychologica, 137*, 318–329.

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M. M., Turner, R., & Ungerleider, L. G. (1998). The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. *Proceedings of the National Academy of Sciences, 95*(3), 861–868.

Korman, M., Raz, N., Flash, T., & Karni, A. (2003). Multiple shifts in the representation of a motor sequence during the acquisition of skilled performance. *Proceedings of the National Academy of Sciences of the USA. 100*, 12492–12497.

Mattoon, J. S. (1994). Designing instructional simulations: Effects of instructional control and type of training task on developing display-interpretation skills. *International Journal of Aviation Psychology, 4*, 189–209.

Muller, B. (1999). Use specificity of cognitive skills: Evidence for production rules? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 191–207.

Naylor, G., & Briggs, J. (1963). Effects of task complexity and task organization on the relative efficiency of part and whole training methods. *Journal of Experimental Psychology, 65*, 217–224.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*, 1–32.

Park, J. H., Wilde, H., & Shea, C. H. (2004). Part-whole practice of movement sequences. *Journal of Motor Behavior, 36*, 51–61.

Pascual-Leone, A., Grafman, J., Clark, K., Stewart, M., Massaquoi, S., Lou, J.-S., & Hallett, M. (1993). Procedural learning in Parkinson's disease and cerebellar degeneration. *Annual Neurology, 34*, 594–602.

Rand, M. K., Hikosaka, O., Miyachi, S., Lu X., Nakamura, K., Kitaguchi, K., & Shimo, Y. (2000). Characteristics of sequential movements during early learning periods in monkeys. *Experimental Brain Research, 131*, 293–304.

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 585–594.

Rozanov, S., Keren, O., & Karni, A. (2010). The specificity of memory for a highly trained finger movement sequence: Change the ending, change all. *Brain Research, 1331*, 80–87.

Ruitenberg, M. F. L., Abrahamse, E. L., De Kleine, E., & Verwey, W. B. (2012a). Context-dependent motor skill: Perceptual processing in memory-based sequence production. *Experimental Brain Research, 222*, 31–40.

Ruitenberg, M. F. L., De Kleine, E., Van der Lubbe, R. H., Verwey, W. B., & Abrahamse, E. L. (2012b). Context-dependent motor skill and the role of practice. *Psychological Research, 76*, 812–820.

Ruitenberg, M. F. L., Verwey, W. B., & Abrahamse, E. L. (2015). What determines the impact of context on sequential action? *Human Movement Science, 40*, 298–314.

Schmidt, R. A., & Wrisberg, C. A. (2008). *Motor learning and performance: a problem-based learning approach* (4th edn.). Champaign: Human Kinetics.

Sosnik, R., Hauptmann, B., Karni, A., & Flash, T. (2004). When practice leads to co-articulation: The evolution of geometrically defined movement primitives. *Experimental Brain Research, 156*, 422–438.

Speelman, C. P., & Kirsner, K. (2001). Predicting transfer from training performance. *Acta Psychological, 108*, 247–281.

Vakil, E., Karni, A., & Loewenschuss, I. (2002). Time delay and transfer in different stages of learning the tower of Hanoi. *43rd Annual Meeting of the Psychonomic Society*.

Verwey, W. B. (2003). Effect of sequence length on the execution of familiar keying sequences: Lasting segmentations and preparation? *Journal of Motor and Behavior, 35*, 343–354.

Verwey, W. B. (2003a). Processing modes and parallel processors in producing familiar keying sequences. *Psychological Research, 67*, 106–122.

Verwey, W. B., & Dronkert, Y. (1996). Practicing a structured continuous key-pressing task: Motor chunking or rhythm consolidation? *Journal of Motor Behavior, 28*, 71–79.

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(6), 1047–1060.