It is not uncommon for research into specific memory deficits in certain disabled groups to offer important insights into general memory processes overall (e.g., implicit memory in amnesia; see Schacter, 1987). The current study investigates implicit sequence learning in primary school children and higher education students with dyslexia. As learning to read words and texts can be understood as the acquisition of grapheme and phoneme sequences, studying sequence learning in dyslexics can potentially improve our insights into the mechanisms and impact of sequence learning in general. Developmental dyslexia is commonly defined as a disability characterized by low reading achievement and deficiencies in learning to spell and write (Snowling, 2012). Since the beginning of the research into dyslexia, a number of causal hypotheses have been formulated. The literature is divided into two types of causal hypotheses. On the one hand, most researchers believe in a rather specific deficit. On the other, some researchers seek an underlying cause of dyslexia in the domain of general cognitive processes. Before discussing the literature about implicit sequence learning in dyslexia, we first take a brief look at these diverse explanations of dyslexia.

The most dominant theory, supported by a majority of reading researchers, attributes the specific problems associated with dyslexia to a phonological processing deficit (for reviews, see Stanovich & Siegel, 1994; Vellutino, Fletcher, Snowling, & Scanlon, 2004; Ziegler & Goswami, 2005). These phonological problems cause difficulties in learning grapheme–phoneme associations and in mappings between how words appear in print and how they sound (Bradley & Bryant, 1983; Ramus et al., 2003; Snowling, 2001). However, this phonological deficit is not presumed to be entirely specific to the reading domain. Indeed, a fundamental problem with the phonological deficit hypothesis is why the phonological processing deficit fails to manifest itself or is much less evident outside the domain of the reading process. Indeed, one would expect the extremely complex phonological processes in speech perception and production to be even more vulnerable to a phonological deficit than the phonological processes involved in reading relatively simple words at the start of reading development. Recently, several studies have cast doubt about the precise nature of the presumed dyslexic phonological deficit. It seems that the phonological representations of persons with dyslexia are normal, but the phonological problems they experience can only be demonstrated when task requirements enhance the processing load in terms of short-term memory (STM), conscious awareness, or time constraints (Ramus & Szenkovits, 2008). Hence, the basic problem of individuals with dyslexia is not in the nature of the phonological representations but rather in accessing these representations (see also Boets et al., 2013).

Even though the very hallmark of dyslexia is a specific failure to learn to read, dyslexic persons show impairments on a wider variety of cognitive tasks. Cognitive differences between reading-
disabled and nondisabled persons are found virtually everywhere. It has been demonstrated that poor readers score lower than normal readers on syntactic awareness tasks (Bentin, Deutsch, & Liberman, 1990; Chung, Ho, Chan, Tsang, & Lee, 2013), on measures of general linguistic awareness (Johnson, 1993; Siegel & Ryan, 1984), on general rule learning tasks (Fletcher & Prior, 1990; Manis et al., 1987; Morrison, 1984), on short term memory and processing strategy tasks (Boets, De Smedt, Cleuren, Vandewalle, & Ghesquière, 2010; Brady, 1991; Share, 1994; Siegel, 1994; Torgesen, 1978), and on measures of metacognitive strategies (Baker, 1982; Wong, 1991). These results are consistent with the one-half standard deviation IQ deficit displayed by poor readers (Stanovich, 1986). Based on the observation that dyslexic persons often present a range of impairments, some researchers believe that the underlying cause of dyslexia should be situated in a more general cognitive process.

One example of a deficit in a general cognitive process that might have a negative impact on the acquisition of reading and writing skills, because learning to read involves both explicit and implicit processes, is a deficit in implicit learning. Initially, children learn grapheme–phoneme correspondences explicitly and often under supervision. Afterward, they apply and continue to learn how phonology is mapped into its written representation implicitly and unsupervised (Gombert, 2003; Petersons & Reis, 2006; Ziegler & Goswami, 2005). Yet the literature on implicit learning in dyslexia is relatively new and has produced mixed results (for reviews, see Folia et al., 2008; Orban, Lungu, & Doyon, 2008).

Explicit learning in dyslexic persons has often been investigated by using different variants of the serial reaction time (SRT) task (Nissen & Bullemer, 1987). Although no task is process pure (e.g., Destrebecqz & Cleeremans, 2001), learning in an SRT task is called implicit because the acquisition of the knowledge about the presented sequence occurs without explicit learning instructions and is also observed if the presented sequence is not explicitly discovered by the participant. In this paradigm, participants have to respond as quickly and as accurately as possible to stimuli that follow a sequential pattern. The participants are unaware of the existence of the repeating sequence. Learning is then demonstrated by a decrease in reaction time and number of errors in response to the sequential pattern compared to random stimulus presentation. Several studies have reported that dyslexic children are impaired in implicit sequence learning (Menghini, Hagberg, Caltagirone, Petrotsini, & Vicari, 2006; Stoodley, Harrison, & Stein, 2006; Stoodley, Ray, Jack, & Stein, 2008; Vicari et al., 2005; Vicari, Marotta, Menghini, Molinari, & Petrotsini, 2003). However, other studies failed to find a specific dyslexic deficit in implicit sequence learning (Deroost et al., 2010; Kelly, Griffiths, & Frith, 2002; Rodenrys & Dunn, 2008; Weber et al., 2003).

The focus of the presented study is yet another paradigm that can be used to investigate whether dyslexic students suffer from a sequential learning deficit: the Hebb repetition learning paradigm. Hebb (1961) asked his participants to perform an immediate serial recall task in which one specific sequence of digits was repeated every third trial. Participants performed significantly better in the recall of repeating sequences compared with nonrepeating sequences. This phenomenon was called the Hebb repetition effect. In a recent study, Szmalec, Loncke, Page, and Duyck (2011) demonstrated that the Hebb repetition effect is affected in dyslexic individuals, even for nonverbal modalities. Their interpretation of the results is that dyslexics do not suffer from an overall implicit sequence learning deficit but rather experience difficulties with the consolidation or transfer of serial-order information, initially stored in STM memory, into a stable long-term memory (LTM) trace. Consolidation in LTM is seen as a partly independent or separate modality-nonspecific process (Szmalec et al., 2011, p. 1275). The SRT task differs from the Hebb repetition paradigm in such a way that the Hebb effect is not really a measure of pure implicit learning, because participants are mostly aware of the repeating sequence. However, awareness appears to have no impact on the degree of learning in this paradigm (e.g., Stadler, 1993; but see Weitz, O’Shea, Zook, & Needham, 2011). In other words, Hebb learning seems to be implicit in some participants but explicit in others.

Importantly, Szmalec et al. (2011) have made a persuasive argument for the crucial role of processing serial order information in language learning. As learning serial order information is commonly believed to play a crucial role in cognition because many aspects of daily human behavior are sequential in nature, this ability is probably very essential in language learning and language processing as well (e.g., Conway & Christiansen, 2001; Page & Norris, 2008, 2009). The precise nature of the relation between serial-order learning and language learning has long been debated (e.g., Conway & Pisoni, 2008). The idea that the Hebb repetition effect is a laboratory analogue of novel word learning has further been elaborated by Szmalec, Duyck, Vandierendonck, Mata, and Page (2009). In their study, sequences of nonsense syllables (e.g., zi-lo-ka-ko-fi-se-be-ra-mo) were used in a Hebb learning procedure. Afterward, an auditory decision task was administered. The nonwords in this task were constructed from the syllables used in the repeating Hebb sequences (zikloka, hofise, beramo). Response times on the Hebb-based nonwords were significantly longer than on the control nonwords—that is, they were more slowly identified as nonwords. Apparently, the repeated sequences of syllables learned in the Hebb procedure had established stable, long-term lexical representations. The researchers concluded that this procedure is similar to the way children implicitly learn words from sequence regularities in the phonological and orthographic input from their environment (see also Mosse & Jarrold, 2008, for converging correlational evidence). These research findings provide the rationale for the hypothesis that the various reading and spelling problems experienced by dyslexic individuals arise from impairments in learning serial order information in LTM. Bogaerts, Szmalec, Bachmann, Page, and Duyck (2013) present this SOLID (serial-order learning impairment in dyslexia) hypothesis as a novel integrative account of dyslexia. The SOLID account of dyslexia can be linked to the double deficit theory of dyslexia (Wolf & Bowers, 1999). According to this theory, processes underlying the ability to rapidly name letters presented serially are a second source of reading problems, apart from a phonological deficit. Interestingly, the serial format of rapid naming (RAN) was found to correlate stronger with reading than the discrete format (Bowers, 1995). The rapid integration of visual and orthographic processes, tapped by RAN tasks, would be crucially required in learning to read (Bowers & Ishaik, 2003). In two experiments we attempted to replicate Szmalec et al.’s (2011) finding of a Hebb learning deficit in dyslexics, including some methodological improvements. Before presenting the experiments, we discuss these methodological issues.
Methodological Issues and Improvements

The first issue concerns the well-established comorbidity of developmental dyslexia and attention deficit disorders (ADD; Araujo, 2012; Boada, Willcutt, & Pennington, 2012). Because the Hebb learning task is very demanding on sustained and focused attention, a differential Hebb effect is not necessarily the result of a deficit in serial learning but may be attributed to the possible comorbid attention problems of persons with dyslexia (see also Wimmer’s analogous critique on the automatization deficit hypothesis of Nicolson & Fawcett, 1990; Wimmer, Mayringer, & Raberger, 1999). Because this comorbidity does not imply that each individual with dyslexia suffers from (minor) attentional problems—indeed, many dyslexics do not show any attentional problems at all—it is essential to avoid this confound. The most sensitive way to control for this confounding factor without losing statistical power is to analyze the complete data once including and once without including attentional functioning as a covariate.

Another even more important methodological issue concerns the statistical analysis performed by Szmalec et al. (2011). The slopes of the regression lines based on performances on successive Hebb and filler sequences for each individual participant were used as the dependent variable in an analysis of variance (ANOVA). One serious problem with this method is that slopes and intercepts are often negatively correlated, especially when maximum performance has an upper bound. If the intercept (initial performance) is already high, there is little room for improvement. This means that the slope of a person who starts his sequence with a very good score tends to be lower than the slope of a person who starts his sequence with a very low score. In fact, the mean correlation between the intercepts and slopes for the three Hebb conditions in the Szmalec et al. (2011) study was substantial (r = −.54, p < .01), as was also the case in our replication study (r = −.52, p < .01). This means that slopes are poor measures and should be avoided, because they inevitably miss a substantial portion of crucial variance in Hebb learning. In order to adequately capture the gradualness of Hebb learning over consecutive trials, we constructed a new index based on a weighted sum of correct responses.

A third methodological improvement involves the use of state trace analysis as a matching technique (Van den Broeck & Geudens, 2012). This method allows a direct comparison of Hebb trial performance when performance on filler trials is equated between the groups. Therefore, this method is more sensitive to detect a specific Hebb learning deficit because by matching dyslexic and control subjects on filler performance, both groups are matched on sequential learning without the crucial phase of consolidation in LTM. This method of equating performances on filler trials is also necessary bearing in mind the often reported verbal STM problems of poor readers. Indeed, if the aim of the study is to compare performances on Hebb sequences with performances on filler sequences, it is mandatory that performances on filler trials be equated between groups.

A final methodological issue pertains to the scoring method that Szmalec et al. (2011) used to evaluate Hebb learning. In this method an item was scored as correct if and only if it was recalled in its correct position. Hence, items that were recalled in a correct serial order but in the wrong position were not counted. This means, for example, that if a participant recalled all items of a sequence in the correct order except the first one, his or her score would be 0. So in fact, only the position of the items is considered, not the serial order. We address this problem by using the McKelvie scoring method (McKelvie, 1987). In this method, both the position and serial order of items are taken into account.

Experiment 1

In this first experiment Hebb learning was examined in adolescents by replicating the study by Szmalec et al. (2011). The results were analyzed once using exactly the same method and data-analytical procedures as applied in the Szmalec et al. (2011) study, and once taking into account some methodological improvements. Potential methodological enhancements were the use of a statistical control on attentional processes, the procedure of weighted sum scores to detect Hebb learning, the method of state trace analysis as an improved matching technique, and McKelvie’s scoring method (McKelvie, 1987). Except for state trace analysis, which has already been proven to offer a better match on baseline performance, each of these adaptations was tested empirically before they were adopted in the data analysis.

Method

Participants. A total of 59 higher education students participated in this study. Twenty-six students (17 females, nine males) had an official diagnosis of dyslexia, and 33 students (25 females, eight males) were IQ-matched control students without any reading problems. Dyslexic participants were diagnosed either by an individual speech therapist or by a specialized center. The diagnoses were all based on three criteria that are used by the Stichting Dyslexie Nederland (Netherlands Dyslexia Foundation; 2008): (a) reading and/or spelling abilities significantly below the level of performance expected for their age (i.e., below the 10th percentile); (b) resistance to instruction despite effective teaching; (c) impairment that cannot be explained by extraneous factors, such as sensory deficits. For further validation, two Dutch reading tests that are diagnostic for dyslexia were administered. The first was the One Minute Test (OMT; Brus & Voeten, 1973), a word reading test in which participants are instructed to correctly read aloud as many words as possible. The raw score is the number of correctly read words after 1 min of reading. The second was the Klepel (Van den Bos, Lutje Spelberg, Scheepmans, & de Vries, 1994), a non-word reading test in which participants are instructed to correctly read aloud as many nonwords as possible within 2 min.

To match the dyslexic and the control group on IQ, a short-form IQ measure was used (Turner 1997), including the Similarities, Comprehension, Block Design and Picture Completion subtests of the Wechsler Adult Intelligence Scale (Dutch version; 3rd ed.; Wechsler, 1998). Our experimental procedure also included a test of sustained attention and two questionnaires concerning attentional functioning. The test of sustained attention was the Bourdon-Wiersma (van der Ven & Smid, 1989), a dot cancelation test. The test consisted of a scoring form with 50 lines, each line containing 25 groups of dots (each group consists of three, four, or five dots in random order). Participants were instructed to cross out all groups of four dots (targets). Each line contained eight targets in a random order. The experimenter used a digital stopwatch to measure the time participants required for each line. Afterward, the number of correctly marked groups of four dots was determined.
and the number of correct marked groups of four dots per second was calculated. Two versions of a Dutch self-report questionnaire (ZVAH; Baeyens, Van Dyck, Broothaerts, Danckaerts, & Kooij, 2012) to detect attention problems were used. Both versions contained the same questions. The first one consisted of questions regarding the participants’ childhood; the second asked about current attentional functioning. As can be seen in Table 1, the group of dyslexic students and the control group differed significantly only on the two measures that are diagnostic for dyslexia, on the test of sustained attention and on the self-report questionnaire concerning attentional problems during childhood. These measures clearly indicate that the clinically diagnosed dyslexic group also suffered from (mild) attentional problems.

**Design, procedure, and materials.** Testing took place on an individual basis in a quiet room at the participant’s school or home. The experimental procedure consisted of two test phases. Each test phase lasted approximately 1 hr. To prevent experimenter bias, test assistants were extensively instructed to consider all collected data as potentially interesting and to assume an empirical attitude that sets each theoretical position on an equal footing. This is a standard procedure at our lab.

**First phase.** During the first testing phase the two reading tests, the four IQ subtests of the WAIS–III, the attention test, and the attention questionnaires were administered.

**Second phase: Hebb tasks.** In the second phase, the Hebb procedure was carried out (Hebb, 1961). In this procedure, participants are asked to perform an immediate serial-recall task in which one particular sequence is repeated every third trial. The Hebb procedure we used was similar to that of Szmalec et al. (2011) and was programmed in E-Prime (Version 1.2; Schneider, Eschman, & Zuccolotto, 2002). The order of the three Hebb tasks was counterbalanced across participants. We assessed awareness of the repetition of Hebb sequences in a postexperimental interview (after the three Hebb tasks) that consisted of one question: “What did you notice about the different sequences presented in the three tasks?” For each task the answer of the participant was scored as “yes” or “no” according to their awareness of the sequences.

**Verbal–visual Hebb condition.** We adopted the two lists of nine nonsense syllables constructed by Szmalec et al. (2011): da-fi-ke-mo-pu-sa-ti-vo-zu and ba-du-ki-le-mu-so-to-vi-za. For each individual, one list was used in the verbal–visual Hebb task, the other in the verbal–auditory Hebb task. The use of both lists in one of the two presentation modalities was counterbalanced across participants. Thirty sequences of nine nonsense syllables were presented serially in the middle of a computer screen to the participants for immediate serial recall. The 30 sequences consisted of 20 nonrepeated sequences (filler sequences) and one Hebb sequence that was repeated 10 times. Participants were not informed that one sequence (the Hebb sequence) recurred every third trial. At recall, all nine syllables and one question mark were distributed randomly on the computer screen. Participants were instructed to repeat the nine syllables in the correct serial order. Participants were instructed to say “blank” for every omitted item. The spoken output of the participants was written down by the experimenter. According to our general strategy to maximize the probability for Hebb learning to take place, we considered a potential stimulus-specific order-effect as less important than reducing random error variance. Therefore, the same Hebb sequence and filler sequences were used in the same order for all participants receiving the same list of nonsense syllables (half of the subjects).

**Verbal–auditory Hebb condition.** In the verbal–auditory Hebb condition, 30 sequences of nonsense syllables were presented auditorily to the participants for immediate serial recall. The syllables were digitally recorded by a female speaker and presented auditorily through speakers at a rate of one per second. Thirty sequences consisted of 20 nonrepeated sequences (filler sequences) and one Hebb sequence that was repeated 10 times. Participants were not informed that one sequence (the Hebb sequence) recurred every third trial. As in the previous condition, for all participants receiving the same list of nonsense syllables, the same Hebb sequence and filler sequences were presented in the same order. At recall, the participants were instructed to repeat the nine syllables in the correct serial order. Participants were instructed to say “blank” for every omitted syllable. The spoken output of the participants was written down by the experimenter.

**Visuospatial Hebb condition.** In the visuospatial Hebb condition, we used the dots task, a Corsi-like (Corsi, 1972) visuospatial immediate serial recall task. We used sequences of nine black dots randomly presented on a white background on a computer screen. The dots were presented serially at a rate of one per second, with each dot on a different location on the screen. As Hebb learning in the Szmalec et al. (2011) study was particularly weak in the visuospatial condition, we chose a relatively easy dot configuration

### Table 1

**Characteristics of the Experimental and Control Group**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Control (n = 33)</th>
<th>Dyslexic (n = 26)</th>
<th>Group difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in years</td>
<td>M SD</td>
<td>M SD</td>
<td>t p</td>
</tr>
<tr>
<td>Word reading test</td>
<td>21.64 1.83</td>
<td>20.69 2.45</td>
<td>1.69 .10</td>
</tr>
<tr>
<td>Nonword reading test</td>
<td>88.36 13.82</td>
<td>70.73 11.93</td>
<td>5.16 .00</td>
</tr>
<tr>
<td>WAIS-III Picture Completion</td>
<td>97.24 12.25</td>
<td>70.96 15.58</td>
<td>7.26 .00</td>
</tr>
<tr>
<td>WAIS-III Similarities</td>
<td>11.21 2.16</td>
<td>11.92 2.04</td>
<td>1.29 .20</td>
</tr>
<tr>
<td>WAIS-III Block Design</td>
<td>13.00 1.95</td>
<td>13.23 2.01</td>
<td>0.45 .66</td>
</tr>
<tr>
<td>WAIS-III Comprehension</td>
<td>10.55 2.53</td>
<td>10.77 2.18</td>
<td>0.36 .72</td>
</tr>
<tr>
<td>WAIS-III Comprehension</td>
<td>13.15 2.21</td>
<td>13.42 2.06</td>
<td>0.48 .63</td>
</tr>
<tr>
<td>Bourdon-Wiersma no. correct items/second</td>
<td>0.648 0.08</td>
<td>0.581 0.11</td>
<td>2.65 .01</td>
</tr>
<tr>
<td>Self-report questionnaire; attention childhood</td>
<td>12.00 2.92</td>
<td>13.54 2.67</td>
<td>2.09 .04</td>
</tr>
<tr>
<td>Self-report questionnaire; attention present</td>
<td>10.91 3.12</td>
<td>12.27 2.93</td>
<td>1.71 .09</td>
</tr>
</tbody>
</table>
to enhance the probability to observe Hebb learning. Dot locations and sequences were the same for all participants. As in the two previous conditions, 20 filler sequences were nonrepeated sequences, and one Hebb sequence was repeated at every third trial. After each sequence, all dots were shown on the screen, and the participant used the mouse to indicate the correct order in which the dots had occurred.

Scoring methods and empirical tests of sensitivity. Two scoring methods were adopted. First, we used the same scoring method as Szmalec et al. (2011). In this scoring method, for every recalled sequence a score was defined by counting every recalled item in the same position as that in which it was initially presented. This means that items that were recalled in a correct serial order but in the wrong position were not counted. To address this problem, we also used McKelvie’s scoring method. In this method, position and serial order of items are taken into account (for details, see McKelvie, 1987).

After adopting one of the two scoring methods, a measure of Hebb learning over the 10 trials was required. As already explained, using the slopes of the learning curves is a dubious method to detect Hebb learning. For that reason we constructed a new index of Hebb learning that avoids the disadvantages of using slopes. Of course, one could simply take the mean or summed Hebb performance over the 10 trials, but by thus weighting each trial equally, this method seems to miss the fact that learning improves gradually over trials. To capture this idea of gradualness and to take into account the empirical learning curves, we calculated a weighted sum for each individual by weighting the Hebb score for each of the 10 Hebb trials, based on the predicted overall difference between the filler score of a given series and the Hebb score of the corresponding series. More specifically, for each condition, we took the mean Hebb performance over participants for each of the 10 Hebb trials and the mean filler performance of each of the two filler trials preceding a Hebb trial, and then calculated a linear or quadratic regression equation (whichever fitted the data better) for the filler trials and for the Hebb trials. Then, the differences between the predicted filler scores and Hebb scores served as the weights for calculating the weighted sum.1

To determine which scoring method and index of Hebb learning were to be used in our improved data analysis, we tested these methods empirically on their criterion validity by checking how well they correlated with the reading scores and with being dyslexic or not. From Table 2 it emerges that the method of using slopes is far inferior to the method of the weighted sums, because all correlations with the reading measures and with the diagnostic group variable are considerably larger for the latter. As the correlations of the method of weighted sums using the McKelvie scoring system were very similar and certainly not larger than those of the classic scoring method, we proceed in our improved data analysis with the method of weighted sums based on the Szmalec scoring method. It is noteworthy that the substantial correlations between the improved Hebb learning index and diagnostic category cannot be taken as evidence for a specific Hebb learning deficit in dyslexia, as no attempt is made here to control for filler performance.

In a final preliminary analysis we tested whether attentional functioning could potentially be a confounding factor by inspecting the correlations of the three tests on attentional functioning with the three measures of Hebb learning (the weighted sums for the three conditions). Three out of nine correlations were statistically significant (ranging between .25 and .35). Although attentional functioning does not seem to play a major role in Hebb learning, given that attentional functioning is also correlated with diagnostic category (see Table 1), it seems warranted to analyze the data once with and once without a statistical control for attentional functioning.

Results

First, we present the mean filler and Hebb scores for the typical and for the dyslexic readers over the three conditions, and compare them with those of the Szmalec et al. (2011) study (see Table 3).2 We observed a clear Hebb learning effect in all three conditions for control subjects as well as for the dyslexic participants. The magnitude of this Hebb effect seems to be similar to the Szmalec study, at least for the verbal–visual and for the verbal–auditory conditions. For the visuospatial condition, however, the picture is clearly different. In this case the Hebb effect in the Szmalec study is clearly smaller for the control subjects than in our study, and even nonexistent for the dyslexic subjects. Another notable observation is that although our dyslexic participants scored quite similar on the filler trials compared to their compereers in the Szmalec study, they performed significantly inferior on these trials compared to the normal readers in our study. The difference between the normal and dyslexic readers in the Szmalec study on the filler trials was much smaller and even nonexistent in the visuospatial

---

1 For instance, the regression equation for filler trials in the verbal–auditory condition was F-score = 3.64 + .08Trial, and for the Hebb trials H-score = 4.26 + .31Trial. By taking the difference of both equations, the weights-equation resulted: W = 62 + .23Trial. Because Hebb learning cannot have taken place on the first Hebb trial, this weight was fixed to zero. To obtain the weighted sum for each individual, the Hebb score on each trial was multiplied by the appropriate weight and then summed.

2 Mean Hebb scores proved to be as sensitive to detect Hebb effects compared to the weighted sum scores, as was shown by the correlations with diagnostic category and the reading scores.
F. (2011), however, the crucial Group and modality (verbal–visual vs. verbal–auditory vs. visuospatial) subjects independent variable, and sequence type (filler vs. Hebb) gradient values were entered into a multivariate analysis of variation:

**Table 3**

<table>
<thead>
<tr>
<th>Group and condition</th>
<th>M</th>
<th>SE</th>
<th>M</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal readers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal-visual</td>
<td>4.44</td>
<td>.24</td>
<td>6.21</td>
<td>.28</td>
</tr>
<tr>
<td>Verbal-auditory</td>
<td>4.35</td>
<td>.26</td>
<td>6.47</td>
<td>.25</td>
</tr>
<tr>
<td>Visuospatial</td>
<td>3.93</td>
<td>.15</td>
<td>6.58</td>
<td>.25</td>
</tr>
<tr>
<td>Dyslexic readers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal-visual</td>
<td>3.38</td>
<td>.24</td>
<td>5.16</td>
<td>.27</td>
</tr>
<tr>
<td>Verbal-auditory</td>
<td>2.74</td>
<td>.20</td>
<td>4.44</td>
<td>.36</td>
</tr>
<tr>
<td>Visuospatial</td>
<td>3.63</td>
<td>.15</td>
<td>5.66</td>
<td>3.30</td>
</tr>
</tbody>
</table>

Effect sizes between groups ($d^*$)

- Verbal-visual: 0.82
- Verbal-auditory: 1.26
- Visuospatial: 0.37

Note. For comparability the presented mean filler and Hebb scores are simple arithmetic means and not weighted sum scores.

condition. Figure 1 shows the mean number of correctly recalled Hebb and filler items as a function of trial number, group and presentation modality.

In a first analysis we tested the hypothesis that dyslexic students show impaired Hebb sequence learning compared to age- and IQ-matched control students. Here, we analyze our data exactly as Szmalec et al. (2011) did in their study. In this analysis, scores are calculated using the classic scoring method. An item was scored as correct if it was recalled in its correct serial position. To measure the Hebb repetition effect, for each participant the gradient of the regression line through points representing the performance on Hebb repetitions was compared with the gradient of the regression line for corresponding filler items (e.g., Page, Cumming, Norris, Hitch, & McNeil, 2006; Szmalec et al., 2011). The gradient values were entered into a multivariate analysis of variance (MANOVA) with group (control vs. dyslexic) as a between-subjects independent variable, and sequence type (filler vs. Hebb) and modality (verbal–visual vs. verbal–auditory vs. visuospatial) as within-subject independent variables. We observed significant main effects of sequence type, $F(1, 57) = 126.13, \eta^2_p = .689, p < .000$; of modality, $F(2, 56) = 4.586, \eta^2_p = .141, p = .014$; and of group, $F(1, 57) = 4.869, \eta^2_p = .079, p = .031$. Unlike Szmalec et al. (2011), however, the crucial Group × Sequence type interaction effect was not significant, $F(1, 57) = .128, \eta^2_p = .002, p = .722$, indicating a similar Hebb effect for the control and the dyslexic group. Planned comparisons of the Group × Sequence type interaction effect for the three modalities confirmed the absence of a differential Hebb effect (for the verbal–visual condition: $F(1, 57) = 0.143, \eta^2_p = .003, p = .707$; for the verbal–auditory condition: $F(1, 57) = 1.355, \eta^2_p = .023, p = .249$; for the visuospatial condition: $F(1, 57) = 0.0003, \eta^2_p = .000, p = .987$).

In our second analysis, the previously explained methodological weaknesses are addressed. First, a weighted sum was used as an improved index of Hebb learning and state trace analysis was adopted to be able to compare directly the performance on Hebb trials to the performance on filler trials. Second, the analyses were performed once without statistically controlling for attentional functioning, and once with this control.

In state trace analysis (STA), performance on Hebb sequences is regressed separately on performance on filler sequences for the dyslexic group and the control group. We tested whether a single line is suitable to explain the data (the null model not including reading group) or whether two different lines (one for dyslexic students and one for control students) are needed to describe the relation between Hebb sequence performance and filler sequence performance (the full model). If a single line fitted the data, this would imply that the relation between Hebb sequence performance and filler sequence performance is not affected by dyslexia. If, on the other hand, two lines fitted our data better, and the one for dyslexic students were situated lower than the one for control students, this would be direct evidence for a specific Hebb learning deficit in dyslexic students.

In this analysis, the sum of the filler scores and the weighted sum of the Hebb scores were plotted against each other for each participant (see Figure 2). At each performance level on filler trials both groups are equal and can be compared directly on Hebb performance. First, we removed the bivariate outliers based on the Mahalanobis distances by excluding the data-points with distances ($D^2$) above a cutoff value corresponding to the 97.5% quantile of the chi-square distribution (Rousseeuw & Van Zomeren, 1990). This procedure resulted in the removal of one outlier in the bivari–visual condition (a subject of the control group), two outliers in the visuospatial condition (one dyslexic subject and one of the control group). Importantly, in order to validly compare Hebb performance between groups, STA requires the data points to overlap on the filler scores (Prince, Brown, & Heathcote, 2012; Van den Broeck, Geudens, & van den Bos, 2010); otherwise participants in one group are compared with statistically extrapolated but nonexistent participants in another group, resulting in meaningless conclusions. For example, when no dyslexic readers are found in the upper tail of the distribution of
filler scores, the normal readers falling in that region are compared with statistically predicted scores based on the performance of dyslexic readers with lower filler performance. To avoid this problem, the statistical analyses are exclusively based on the overlapping part of the filler distributions. This procedure resulted in the removal of nine additional subjects in the verbal–visual condition (three dyslexic and six control subjects), 13 subjects in the verbal–auditory condition (one dyslexic and 12 control subjects), and 11 subjects in the visuospatial condition (four dyslexics and seven control subjects). Although no mechanical procedure exists to determine the overlapping distributions, in the case of the verbal–auditory and visuospatial conditions the overlapping regions on the filler scores were clearly discernable. Only for the verbal–visual condition it was less clear which data points to retain and which to remove. However, as we report forthwith, the statistical conclusions for this condition were quite similar if we analyzed the data without removing any subject.

Then, we tested in a hierarchical regression analysis whether group contributed significantly to Hebb performance after including filler performance in the regression equation. This analysis

Figure 1. Mean number of correctly recalled filler and Hebb items in the different conditions, for both typical and dyslexic readers. Accuracy values for filler trials represent the average of the two filler trials preceding each Hebb repetition.
Figure 2. A: State-trace plot for the verbal–visual condition for typical readers and dyslexic readers. B: State-trace plot for the verbal–auditory condition for typical readers and dyslexic readers. C: State-trace plot for the visuospatial condition for typical readers and dyslexic readers.
showed that adding group as a predictor did not significantly improve fit in the verbal–visual task ($R^2$ null model = .340; $\Delta R^2 = .001$; $F$ change $[1,46] = .103; p = .750$) and in the visuospatial task ($R^2$ null model = .121; $\Delta R^2 = .002$; $F$ change $[1,43] = .113; p = .738$). In the verbal–auditory task a significant group effect was found ($R^2$ null model = .464; $\Delta R^2 = .050$; $F$ change $[1,42] = 4.280; p = .045$). Hence, the null hypothesis—that the state trace curves for the typical and for the disabled readers do not differ—could not be rejected for two out of three conditions. This means that if there was in reality no difference between the curves of the two groups (if $H_0$ is true), the probability of finding a difference as large as or even larger than in our sample is .75 for the verbal–visual task, .04 for the verbal–auditory task, and .74 for the visuospatial task. As STA entails that the null hypothesis is in fact a substantive hypothesis, these numbers do not seem really convincing. Note that a nonsignificant result due to a lack of power must not be confused with support for the null hypothesis. What we really want to know is the probability that the null hypothesis or the alternative hypothesis is true given the observed data (the inverse probability). To this end, a Bayesian analysis was performed in which the probability of the null model was compared to the probability of the full model given the data and given the assumption that no model was preferred above the other. For this comparison the Bayesian information criterion (BIC) was calculated for both models. The BIC has been proposed by Raftery (1996) as an index to assess the overall fit of a model and allows a comparison of models (see also Long, 1997). Given that BIC assesses whether the model fits the data sufficiently well to justify the number of parameters that are used, the model with the lowest BIC is the best fitting, yet parsimonious model. The BIC values indicated that the null model fitted the data best for two tasks ($\text{BIC}_{\text{null}} = 468.15$ and $\text{BIC}_{\text{full}} = 471.93$) for the verbal–visual task, $\text{BIC}_{\text{null}} = 434.77$ and $\text{BIC}_{\text{full}} = 434.21$ for the verbal–auditory task, $\text{BIC}_{\text{null}} = 448.82$ and $\text{BIC}_{\text{full}} = 452.53$ for the visuospatial task). Based on the differences between these BIC values, the Bayesian factors could be calculated (Kass & Raftery, 1995). The Bayesian factor is equal to the posterior odds in favor of the most likely hypothesis. The posterior odds of the null model ($M_N$) relative to the full model ($M_F$) equal

$$\text{Pr}(M_F / \text{Observed Data}) / \text{Pr}(M_N / \text{Observed Data}).$$

The Bayesian factors favoring the null model were 3.78 and 3.71 for the verbal–visual and the visuospatial tasks, respectively, whereas the Bayesian factor (.56) favored the full model for the verbal–auditory task (the Bayesian factor for the verbal–visual task when all data points were included, including the nonoverlapping, was 2.88 favoring the null model). According to the criteria proposed by Kass and Raftery (1995), the data provided substantial evidence for the null hypothesis for the verbal–visual task and for the visuospatial task, whereas the data for the verbal–auditory task gave barely more support for the full hypothesis than for the null hypothesis. Next we tested whether the previous conclusions were affected by including the variables indicative of attentional functioning into the first step of the hierarchical regression analysis. As the statistical tests showed the same pattern of results—the visual-verbal and visuospatial conditions showing no significant group effect while the verbal–auditory condition did show a significant between-group difference—we only present here the Bayesian factors. The Bayesian factors favoring the null model were 3.71 and 3.18 for the verbal–visual and the visuospatial tasks, respectively. Although the Bayesian factor favoring the full model for the verbal–auditory task was somewhat higher than without controlling for attentional functioning (1.64), this support is considered weak according to Kass and Raftery (1995).

As Szmalec et al.’s (2011) hypothesis is explicitly formulated as modality nonspecific, a more powerful analysis of their hypothesis is possible by integrating all data from the three tasks into one multilevel analysis. As a result of the nested nature of the design (within-person conditions are nested under the between-persons level), within-person variance and between-persons variance should be modeled appropriately. Multilevel modeling (Raudenbush & Bryk, 2002) was used to obtain state trace functions for each individual, to obtain average parameters (and their variances) across participants, and to compare the parameters from these functions in the reading disabled and typical reading samples. This analytic approach first required the specification of a condition-level model (within subject), which represents the performance for each Hebb condition as a function of the performance from the corresponding filler condition for each individual. Second, person-level models (between subjects) were specified representing each regression parameter (intercept and slope) from the condition-level model as a dependent variable. In the null model these parameters were written as a function of overall means and individual unique effects. Additionally, group effects on the regression parameters were included in the full model.

Filler scores and Hebb scores were first transformed to $z$ scores to make them comparable over conditions. Next, we removed the bivariate outliers based on the Mahalanobis distances by excluding the data points with distances ($D^2$) above a cutoff value corresponding to the 97.5% quantile of the chi-square distribution (Rousseeuw & Van Zomeren, 1990). This procedure resulted in the removal of five data points. Additionally, 25 data points, lying outside the overlapping region, were removed. As a consequence, 147 out of 177 ($3 \times 59$) data points were included in the analysis. Fitting the full model using MPlus 7 (Muthén & Muthén, 1998/2012) produced the following parameter estimates: For the average intercept of the typical readers, $\gamma_{00} = .108$ ($SE = .093; p = .244$); for the average slope of the typical readers, $\gamma_{10} = .631$ ($SE = .123; p < .000$); for the difference in intercepts between the reading groups, $\gamma_{01} = -.176$ ($SE = .0142; p = .214$); and for the difference in slopes between the reading groups, $\gamma_{11} = .092$ ($SE = .175; p = .601$). These results show that disabled readers demonstrated a slight but nonsignificant Hebb learning deficit (a lower intercept). However, the null hypothesis—that the state trace curves for the typical and for the disabled readers do not differ—could not be rejected. As evidenced by a likelihood ratio test, the full model was not significantly different from the null model, $\chi^2(2) = 2.152, p = .341$. The BIC value for the null model was 351.10 and for the full model 358.93, resulting in a Bayesian factor favoring the null model of 7.83. This implies that the null model was about 8 times as likely as the full model, a number that can be interpreted as substantial evidence (Kass & Raftery, 1995). The parameter estimate for the average intercept of this null model was $\gamma_{00} = .221$ ($SE = .069; p = .768$), and for the average slope, $\gamma_{10} = .704$ ($SE = .087; p < .000$). Thus, for all participants an increase in performance of one filler item, expressed in $z$ scores,
was associated with an increase in performance of .71 Hebb items. In conclusion, these results show that no different Hebb learning was observed between the two groups across the three modalities.

Nearly 80% of the participants indicated that they were aware of the repeating Hebb sequence (79.7% in the verbal–visual condition, 78.0% in the verbal–auditory condition, and 79.9% in the visuospatial condition). However, participants of the control group detected the Hebb sequences more frequently than the dyslexics (about 89% vs. 65%). Importantly, Hebb performance was better when the sequence was detected, as indicated by the positive biserial correlations between mean Hebb score performance and awareness: $r = .26, p = .045$ for the verbal–visual condition, $r = .30, p = .022$ for the verbal–auditory condition, and $r = .52, p < .000$ for the visuospatial condition. Nevertheless, the Hebb learning effect was still present for the participants who were unable to detect the sequences (the main effect of sequence type was statistically significant for each modality, all $p$s < .02). Finally, as an alternative test of a differential Hebb effect, we tested if reading scores, as a continuous measure of reading ability, were correlated with Hebb scores after the influence of the filler-scores was partialed out from the Hebb-scores: All $p$s were nonsignificant.

### Discussion

The purpose of Experiment 1 was to try to replicate a differential Hebb-learning effect in disabled versus normal adolescent readers, as reported by Szmalec et al. (2011). First, as our results showed, we observed a clear Hebb-learning effect in all three conditions for control subjects as well as for the dyslexic participants, implying that Hebb learning was confined neither to specific stimulus material nor to specific order effects. Thus, it appears that our experimental procedures were more sensitive in detecting Hebb learning than those in the Szmalec study, since in the latter the Hebb effect in the visuospatial condition was smaller for the normal readers and even nonexistent in the dyslexic readers (see Table 3). This last observation is atypical, since filler performance in their visuospatial condition was even somewhat better for the dyslexics than for the normal readers. Nevertheless, we could not identify a specific Hebb learning deficit in dyslexics in any of the three modalities (verbal–visual, verbal–auditory, and visuospatial), when applying the data-analytical procedure used by Szmalec et al. However, we demonstrated empirically that the use of slopes is a poor technique to detect Hebb learning and might be the reason why we couldn’t replicate a dyslexic Hebb learning deficit. Therefore, we developed a more sensitive method, the weighted sums, to measure Hebb learning. After adopting this methodological improvement and after equating both groups more precisely on filler performance with STA, we found a significant group effect for the verbal–auditory condition, but not for the two other conditions. Moreover, a Bayesian analysis revealed that even this effect in the verbal–auditory condition could not be interpreted as providing more support for a dyslexic serial order learning impairment than for the hypothesis that no such deficit exists. In the two other conditions the evidence clearly supported the null hypothesis. This conclusion was further strengthened by a multi-level analysis, generating more statistical power, in which the three conditions are appropriately nested under the participant level.

Although our methodological perfections are arguably important (especially STA and the use of a weighted sum as an index of Hebb learning), the reason for failing to replicate the Hebb learning deficit seems not to be related to the method adopted. Moreover, a reanalysis of the data of Szmalec et al. (2011), using STA with mean filler and mean Hebb scores, revealed marginal between-groups effects for the verbal–visual and the verbal–auditory conditions ($p$ values of the $F$-change statistics were .065 and .057, respectively), whereas the visuospatial condition showed a statistically significant effect ($p = .017$). Thus, also in Szmalec’s study, the main culprit seems not to be the method adopted but the data that are genuinely different. We can only speculate on the reasons for these differing results. One plausible possibility is that adolescent or adult dyslexics are not naive subjects; they are reported to sometimes react in a sophisticated or strategic manner. More specifically, Harrison, Edwards, and Parker (2008) provided evidence that postsecondary-level students may be motivated to exaggerate or magnify their symptoms. One could speculate that students in the Szmalec study perceived a direct link between their performance in the experiment and the justification to get compensatory facilities, as these facilities directly depended on the diagnosis given by the diagnostic center involved. In our study, there was no such link at all. One way to solve this problem is to distinguish gross achievement measures from more subtle effects that are little known to the participants (Lindstrom, Coleman, Thomasson, Southall, & Lindstrom, 2011). Another, more obvious way to avoid these complications and probably other indirect effects of carrying a history of reading problems is to replicate the study in more naive subjects (i.e., children). An additional reason for a replication study in children is that, from a theoretical viewpoint, exactly the same predictions should be made for this population. As a matter of fact, such a study can be viewed as a more direct test of the SOLID hypothesis stating that the Hebb repetition effect as a laboratory analogue of novel word learning is disturbed in dyslexic individuals. Although Szmalec et al. (2011) indicated that they were the first to study the Hebb repetition effect in relation to dyslexia, Gould and Glencross (1990) used the Hebb paradigm in children. In their study, the group of disabled readers scored significantly lower on the repeated digits task, but on the repeated blocks task the disabled group performed as well as normal readers. The authors concluded that their data do not support a general deficit in serial organization.

Another possible explanation for the divergent findings is that our dyslexic subjects might be less severely impaired than those in Szmalec et al. (2011) and are merely garden-variety poor readers (Gough & Tunmer, 1986; Stanovich, 1988). This possibility is clearly refuted by the data, as our dyslexics’ intellectual functioning was above average (see Table 1) and their mean word reading score was even lower than that of the dyslexic group in the Szmalec study (70.73 vs. 75.25; the mean nonword reading score was 70.96 in our study vs. 65.44 in theirs). In the same vein, the filler scores of our dyslexic participants were very similar to those of the dyslexics in the Szmalec study. Remarkably however, the filler scores of our dyslexics were clearly inferior to those of the normal readers, whereas in the Szmalec study the difference on filler trials between the normal and dyslexic readers was much smaller and even nonexistent in the visuospatial condition. The pattern in our results could be seen as a confirmation of the alternative hypothesis that dyslexics suffer from a STM serial order learning deficit, rather than a deficit of the transfer from...
STM to LTM (see Hachmann et al., 2014; Perez, Majerus, Mahot, & Poncelet, 2012; but see Staels & Van den Broeck, 2013).

The most plausible explanation for the conflicting findings, however, is that the data of both studies are valid and dependable but that the results of the Szmalec et al. (2011) study on the verbal–visual and verbal–auditory conditions are not in contradiction with the null hypothesis, whereas their visuospatial condition was not sensitive enough to detect Hebb learning in all subjects. We return to this possibility in the final discussion.

**Experiment 2**

In the second experiment, Hebb learning was examined in children by adapting the Szmalec procedure (Szmalec et al., 2011). As in the first experiment, results are analyzed once using exactly the same method and data-analytical procedures as applied in the Szmalec et al. (2011) study, and once taking into account some methodological improvements.

**Method**

**Participants.** Fifty-seven fifth and sixth grade children participated in this study. Their age ranged from 9 years 0 months to 12 years 0 months, with a mean age of 10 years 7 months. Twenty-five children (seven girls, 18 boys) were officially diagnosed with dyslexia; the other 32 IQ-matched children (25 girls, seven boys) never showed any reading problems. As in Experiment 1, dyslexic participants were diagnosed by an individual speech therapist. The diagnoses were all based on three criteria which are used by the Stichting Dyslexie Nederland (Netherlands Dyslexia Foundation; 2008). For further validation two Dutch reading tests that are diagnostic for dyslexia were administered: the One Minute Test (OMT; Brus & Voeten, 1973) and the Klepel (Van den Bos et al., 1994). To match the dyslexic and the control group on IQ, a short-form IQ measure was used (Turner 1997), including the Similarities, Comprehension, Block Design and Picture Completion subtests of the Wechsler Intelligence Scale for Children III (Dutch version; Wechsler, 2005).

We included two tests and one questionnaire to examine attentional functioning. The tests were Score! and Score DT, two subtests of the Test of Everyday Attention for Children (TEA-Ch; Manly, Robertson, Anderson, & Nimmo-Smith, 1999). Both tests evaluated sustained attention. In the Score! subtest, children have to count the number of “scoring” sounds they hear on a tape as if they were keeping the score on a computer game across several trials. In the Score DT subtest, the same attention task of counting scoring sounds is combined with another task involving listening for an animal name that occurs at some stage during a spoken news report. A Dutch ADHD questionnaire (AVL; Scholte & Van der Ploeg, 2005) was used to assess attentional functioning. The questionnaire was completed by the parents and the teacher of the participant. Table 4 shows that the group of dyslexic children and the control group differed on the two measures that are diagnostic for dyslexia and on all attention tests and questionnaires.

**Design, procedure and materials.** Testing took place on an individual basis in a quiet classroom at the participant’s school. The experimental procedure consisted of two test phases. Each test phase lasted approximately 1 hr.

**First phase.** During the first testing phase the two reading tests, the four IQ subtests of the WISC–III, and the attention tests were administered. The two attention questionnaires were completed by the teacher and the parents of each participant.

**Second phase: Hebb tasks.** In the second phase, the Hebb procedure was implemented (Hebb, 1961). The Hebb procedure of Szmalec et al. (2011) was adjusted to our younger participants. To minimize verbal processes, we administered two nonverbal tasks, a visuospatial Corsi blocks task and a visual forms sequences task. As in Gould and Glencross (1990), a digit sequences task was also run. The three Hebb tasks were administered in the same order for all participants. The sequences we used in this study were sequences of only seven items. Similar to our first study, we presented 30 sequences; 20 of them were nonrepeated sequences (filler sequences) and one Hebb sequence was repeated 10 times. For all three Hebb tasks, all participants received the same list of sequences, to reduce random error variance. Participants were not informed that one sequence (the Hebb sequence) recurred every third trial. The children’s awareness of the repetition of Hebb sequences was not assessed. Although the issue of awareness is certainly relevant, it was felt that self-report would not be very reliable in this sample.

**Digit sequences.** For the first Hebb task we used a verbal–auditory task with digit sequences. Thirty sequences of seven digits were presented serially to the participants for immediate

---

**Table 4**

**Characteristics of the Experimental and Control Groups**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Controls (n = 32)</th>
<th>Dyslexics (n = 25)</th>
<th>Group difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>M = 10.41, SD = 0.62</td>
<td>M = 10.76, SD = 0.72</td>
<td>t = 1.99, p = .05</td>
</tr>
<tr>
<td>Word reading test</td>
<td>M = 66.63, SD = 13.72</td>
<td>M = 47.48, SD = 9.76</td>
<td>t = 5.90, p = .00</td>
</tr>
<tr>
<td>Nonword reading test</td>
<td>M = 63.34, SD = 16.71</td>
<td>M = 38.52, SD = 9.38</td>
<td>t = 6.65, p = .00</td>
</tr>
<tr>
<td>WISC-III Picture Completion</td>
<td>M = 8.38, SD = 2.51</td>
<td>M = 8.68, SD = 2.63</td>
<td>t = 0.45, p = .66</td>
</tr>
<tr>
<td>WISC-III Similarities</td>
<td>M = 10.84, SD = 3.18</td>
<td>M = 10.84, SD = 2.94</td>
<td>t = 0.00, p = 1.00</td>
</tr>
<tr>
<td>WISC-III Block Design</td>
<td>M = 8.84, SD = 2.42</td>
<td>M = 8.84, SD = 2.82</td>
<td>t = 0.01, p = 1.00</td>
</tr>
<tr>
<td>WISC-III Comprehension</td>
<td>M = 9.53, SD = 3.55</td>
<td>M = 9.44, SD = 3.31</td>
<td>t = 0.10, p = .92</td>
</tr>
<tr>
<td>TEA-Ch Score!</td>
<td>M = 8.69, SD = 2.22</td>
<td>M = 7.44, SD = 2.16</td>
<td>t = 2.13, p = .04</td>
</tr>
<tr>
<td>TEA-Ch Score DT</td>
<td>M = 10.59, SD = 3.05</td>
<td>M = 8.56, SD = 3.92</td>
<td>t = 2.21, p = .03</td>
</tr>
<tr>
<td>AVL parents (ADHD questionnaire)</td>
<td>M = 3.44, SD = 2.12</td>
<td>M = 4.92, SD = 2.86</td>
<td>t = 2.25, p = .03</td>
</tr>
<tr>
<td>AVL teacher (ADHD questionnaire)</td>
<td>M = 2.78, SD = 2.03</td>
<td>M = 5.04, SD = 2.89</td>
<td>t = 3.46, p = .00</td>
</tr>
</tbody>
</table>
serial recall. The digits were digitally recorded by a female speaker and presented auditorily through speakers at a rate of one per second. At recall, the participants were instructed to repeat the seven digits in the correct serial order. Participants were instructed to say “blank” for every omitted item. The spoken output of the participants was written down by the experimenter. The task was programmed in Microsoft Office PowerPoint 2007.

**Corsi blocks sequences.** The second task was a Corsi block tapping task, a visuospatial task (Corsi, 1972). For this task, seven blocks (3 cm × 3 cm × 3 cm) were randomly glued to a solid piece of cardboard (30 cm × 42 cm; see Figure 3). Participants were instructed to repeat 30 sequences of seven taps on a set of seven blocks that were presented to them. Sequences of blocks were tapped at approximately one block/second. Participants were instructed to say “blank” for every omitted item. The sequence pointed out by the participants was written down by the experimenter.

**Visual form sequences.** The third task was also a visual task. We used sequences of visual forms without any meaning to prevent participants from verbalizing the items (see Figure 4). The visual forms were selected from the Microsoft Word 2007 symbol database. Thirty sequences of seven visual forms were presented serially in the middle of a computer screen to the participants for immediate serial recall. At recall, all seven visual forms were presented randomly on the computer screen. To avoid spatial learning of the response, on each trial another random arrangement of the seven forms was presented. Participants were instructed to point out the seven forms in the correct serial order. Participants were instructed to say “blank” for every omitted visual form. The sequence pointed out by the participants was checked by the experimenter on a sheet of paper with all forms listed. The task was programmed in Microsoft Office PowerPoint 2007.

**Scoring methods and empirical tests of their sensitivity.** As in Experiment 1, the scoring method and index of Hebb learning were empirically tested on their criterion validity by checking how well they correlated with the reading scores and with being dyslexic or not. Although the correlations of Hebb learning with the diagnostic category were smaller in children than in adolescents, as in Experiment 1 the method of using slopes is far superior to the method of the weighted sums because all correlations with the reading measures and with the diagnostic group variable are considerably larger for the weighted sums (some correlations of the slopes are even in the wrong direction; see Table 5). As the correlations of the method of weighted sums using the McKelvie scoring system were very similar and certainly not larger than those of the classic scoring method, we proceed in our improved data analysis with the method of weighted sums based on the Szmalec scoring method.

We tested whether attentional functioning could potentially be a confounding factor by inspecting the correlations of the three tests on attentional functioning with the three measures of Hebb learning (the weighted sums for the three conditions). Three out of nine correlations were statistically significant (ranging between .27 and .36). Although attentional functioning does not seem to play a major role in Hebb learning, given that attentional functioning is also correlated with diagnostic category (see Table 4), it seems warranted to analyze the data once with and once without statistically controlling for attentional functioning.

**Results**

First, we present the mean filler and Hebb scores for the typical and for the dyslexic readers over the three conditions (see Table 6). Although the three tasks differed in degree of difficulty (the visual forms sequences task was clearly the most difficult), Hebb learning appears to have occurred in all three tasks (for the visual forms sequences task a one-tailed paired t test resulted in \( p = .051 \)). Figure 5 shows the mean number of correctly recalled Hebb and filler items as a function of group and presentation modality.

In the first analysis we tested the hypothesis that dyslexic children show impaired Hebb sequence learning compared to age- and IQ-matched children. As in our first experiment, the data are analyzed precisely as in Szmalec et al. (2011). In this analysis, scores are calculated using the first scoring method and slopes are used as an

![Figure 3. Display of Corsi blocks task.](image)

![Figure 4. Visual forms sequences task.](image)

**Table 5**

<table>
<thead>
<tr>
<th>Scoring method</th>
<th>Diagnostic category (dyslexic or control group)</th>
<th>Real word reading (OMT)</th>
<th>Pseudoword reading (Klepel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Szmalec</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope (D)</td>
<td>-.051</td>
<td>-.228</td>
<td>-.127</td>
</tr>
<tr>
<td>Weighted (D)</td>
<td>-.161</td>
<td>.124</td>
<td>.273*</td>
</tr>
<tr>
<td>Slope (B)</td>
<td>-.015</td>
<td>-.074</td>
<td>.014</td>
</tr>
<tr>
<td>Weighted (B)</td>
<td>-.057</td>
<td>.112</td>
<td>.255*</td>
</tr>
<tr>
<td>Slope (VF)</td>
<td>-.055</td>
<td>.118</td>
<td>.200</td>
</tr>
<tr>
<td>Weighted (VF)</td>
<td>-.199</td>
<td>.342**</td>
<td>.428**</td>
</tr>
<tr>
<td>McKelvie</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted (D)</td>
<td>-.139</td>
<td>.143</td>
<td>.075</td>
</tr>
<tr>
<td>Weighted (B)</td>
<td>-.080</td>
<td>.129</td>
<td>.076</td>
</tr>
<tr>
<td>Weighted (VF)</td>
<td>-.137</td>
<td>.346**</td>
<td>.391**</td>
</tr>
</tbody>
</table>

*Note. OMT = One Minute Test; D = digit sequences condition; B = Corsi blocks condition; VF = visual forms condition.*

* \( p < .05. \) ** \( p < .01. \)
index of Hebb learning. The slopes were entered into a MANOVA with group (control vs. dyslexic), sequence type (filler vs. Hebb), and modality (digit sequences, Corsi block sequences, and visual form sequences) as the independent variables. We observed a significant main effect of sequence type, $F(1, 55) = 45.431$, $\eta^2_g = .452$, $p < .000$, but we observed no significant main effect of group, $F(1, 55) = .904$, $\eta^2_g = .016$, $p = .346$, or modality, $F(2, 54) = 2.502$, $\eta^2_g = .085$, $p = .091$. The crucial Group × Sequence type interaction effect was also not significant, $F(1, 55) = .087$, $\eta^2_g = .002$, $p = .769$, indicating a similar Hebb effect for the control and the dyslexic group. Planned comparisons of the Group × Sequence type interaction effect for the three modalities confirmed the absence of a differential Hebb effect (for digit sequences: $F(1, 55) = 0.117$, $\eta^2_g = .002$, $p = .733$; for Corsi block sequences: $F(1, 55) = 0.025$, $\eta^2_g = .000$, $p = .874$; for visual form sequences: $F(1, 55) = 0.537$, $\eta^2_g = .010$, $p = .467$).

In our second analysis, the methodological improvements were included. First, a weighted sum was used as an improved index of Hebb learning and state trace analysis was adopted to be able to compare directly the performance on Hebb trials to the performance on filler trials. Second, the analyses were performed once without statistically controlling for attentional functioning, and once with this control. In these analyses, for each participant, the sum of the filler scores and the weighted sum of the Hebb scores were plotted against each other (see Figure 6).

First, we removed the bivariate outliers based on the Mahalanobis distances by excluding the data-points with distances ($D^2$) above a cutoff value corresponding to the 97.5% quantile of the chi-square distribution. This procedure resulted in the removal of six data points. This time, the data points of both groups nicely overlapped for the digit sequences task ($R^2$ null model = .465; $\Delta R^2 = .000$; $F$ change $[1,52] = .003$; $p = .955$), in the Corsi block sequences task ($R^2$ null model = .388; $\Delta R^2 = .000$; $F$ change $[1,52] = .025$; $p = .875$), or in the visual form sequences task ($R^2$ null model = .291; $\Delta R^2 = .001$; $F$ change $[1,52] = .038$; $p = .845$). As can be inferred from these results, no different Hebb learning was observed between the two groups in all three modalities. As in Experiment 1, a Bayesian analysis was performed in which the probability of the null model was compared to the probability of the full model given the data and given the assumption that no model was preferred above the other. For this comparison, BICs were calculated for both models. The BIC values indicated that the null model fitted the data best for all tasks ($BIC_{null} = 363.73$ and $BIC_{full} = 367.73$ for the digit task, $BIC_{null} = 374.26$ and $BIC_{full} = 378.17$ for the Corsi blocks task, $BIC_{null} = 405.52$ and $BIC_{full} = 409.50$ for the visual form task). The Bayesian factors all favored the null model: 4.00, 3.91, and 3.98 for the digit sequences, the Corsi block sequences and the visual form sequences tasks, respectively. According to Kass and Raftery’s (1995) criteria, the data provided substantial evidence for the null hypothesis for the three tasks.

Next we tested whether the previous conclusions were altered by including the variables indicative of attentional functioning into the first step of the hierarchical regression analysis. As the statistical tests showed the same pattern of results, we only present here the Bayesian factors. The Bayesian factors favoring the null model were 3.93, 3.95, and 3.82 for the digits task, the Corsi blocks task, and the visual forms task, respectively. Clearly, these results were virtually the same as without controlling for attentional functioning.

As in Experiment 1, we integrated all data in one multilevel analysis to obtain more power. Filler scores and Hebb scores were first transformed to $z$ scores to make them comparable over conditions. Next, we removed the bivariate outliers based on the Mahalanobis distances by excluding the data points with distances ($D^2$) above a cutoff value corresponding to the 97.5% quantile of the chi-square distribution. This procedure resulted in the removal of six data points. This time, the data points of both groups nicely overlapped for the filler scores. As a consequence, 165 out of 171 (96.4%) data points were included in the analysis. Fitting a full model, including group as a predictor variable in the person level model, MPlus 7 produced the following parameter estimates: for the average intercept of the typical readers, $\gamma_{00} = .011$ ($SE = .007$; $p = .885$); for the average slope of the typical readers, $\gamma_{10} = .590$ ($SE = .081$; $p = .000$); for the difference in intercepts between the reading groups, $\gamma_{01} = .012$ ($SE = .017$; $p = .92$); and for the difference in slopes between the reading groups, $\gamma_{11} = .022$ ($SE = .137$; $p = .873$). These results show that the null hypothesis—that the state trace curves for the typical and for the disabled readers do not differ—could not be rejected. As evidenced by a likelihood ratio test, the full model was not significantly different from the null model, $\chi^2(2) = 0.04$, $p = .980$. The BIC value for the null model was 389.14 and for the full model 399.31, resulting in a Bayesian factor favoring the null model of 10.17. This implies that the null model was about 10 times as likely as the full model, a number that can be interpreted as substantial evidence (Kass & Raftery, 1995). The parameter estimate for the average intercept of this null model was $\gamma_{00} = .005$ ($SE = .058$; $p = .932$), and for the average slope, $\gamma_{10} = .599$ ($SE = .065$, $p < .000$). Thus, for all participants an increase in performance of one filler item, expressed in $z$ scores, was associated with an increase in performance.
of .60 Hebb items. Summarized, these results show that no different Hebb learning was observed between the two groups across the three modalities. Before discussing these results, we present an aggregated STA that bears on the relationship between the learning of filler sequences and Hebb sequences and their underlying memory mechanisms.

Underlying dimensionality of the memory processes involved. Although we used STA as an equivalence or matching technique (see Van den Broeck & Geudens, 2012) to detect differential Hebb-learning in Experiments 1 and 2, STA as originally developed by Bamber (1979) has recently become an important instrument in detecting underlying causality in cognitive and memory research through the seminal work of Geoffrey Loftus (Loftus, Oberg, & Dillon, 2004; see also Dunn & James, 2003; Newell & Dunn, 2008; Prince et al., 2012). The fundamental question STA seeks to answer is whether the relationship between a set of causal variables and a set of indicator variables is mediated by one or more latent variables (Prince et al., 2012). Szmalec et al. (2011) explicitly stated that “learning in Page and Norris’s model depends both on the quality of the short-term representation of a to-be-recalled list and on an independent weight-change process governed by a variable learning rate” (Szmalec et al., 2011, p. 1275). Hence, consolidation in LTM underlying Hebb performance is partly based on an independent modality-nonspecific process that is different from the STM process needed to remember the filler sequences. The central question is whether the relationship be-

Figure 5. Mean number of correctly recalled filler and Hebb items in the different conditions, for both typical and dyslexic readers. Accuracy values for filler trials represent the average of the two filler trials preceding each Hebb repetition.
Figure 6. A: State-trace plot for the digit sequences condition for typical readers and dyslexic readers. B: State-trace plot for the Corsi blocks sequences condition for typical readers and dyslexic readers. C: State-trace plot for the visual form sequences condition for typical readers and dyslexic readers.
between a set of indicator variables (filler performance and Hebb performance) and a set of causal variables (sequence type and reading group) is determined by the same underlying latent processing mechanism or by a partly independent learning factor over and above serial learning mechanisms in STM (see Figure 7). Note that the figure optionally has a path (dotted arrow) from diagnostic category to STM to capture the possibility that individuals with dyslexia would already have a problem with serial learning in STM. This does not alter the prediction of an additional deficit of order learning in LTM.

The only requirement in STA is the broad assumption that latent variables have a monotonic effect on the indicator variables. Hence, when proficiency at the underlying memory process increases, performance on filler and Hebb sequences should also increase. The specific form this monotonicity increasing (or decreasing) relationship takes for each indicator is irrelevant. From this assumption, a crucial prediction can be derived (Bamber, 1979): If only one latent variable is required to explain the relationship between indicator variables and causal variables, the plot of one indicator against the other must increase (or decrease) monotonically. The outcome of an STA is a state trace plot. The axes of this plot, representing the indicator variables, define a state space for the examined processes (see Figure 8).

The trace factor (modality) in combination with the other causal variables induces variation among the points in the plot, which can be considered “as providing a ‘trace’ of the behavior of the system in the corresponding state space” (Prince et al., 2012, p. 80). If only one latent process underlies the behavior of that system, variation of the latent factor creates a curve that increases (or decreases) monotonically. When we translate this condition to our Hebb learning experiments and hypothesize that the same memory processes are responsible for both Hebb learning in dyslexic and typical readers, all data points should be situated on one monotonically increasing curve. Alternatively, if both reading groups’ underlying memory mechanisms differ, each group will have its own curve. As we have detailed in the analyses of Experiments 1 and 2, this prediction was incorrect. However, by stating that the underlying memory processes are indistinguishable in dyslexic and typical readers, there is still uncertainty about which memory processes are involved. At this point of reasoning, it is possible that for both groups STM processes underlie filler performance, and an additional consolidation process into LTM underlies Hebb performance. From Figure 8, however, it becomes apparent that when the results are averaged out over the participants to reduce unreliability due to individual error variance, filler performance almost perfectly predicts Hebb performance ($r = .998$ in children and $r = .919$ in adolescents). When combining the results of both experiments, we obtained a Pearson correlation of .946 with a bias-corrected and accelerated (BCa) 95% confidence interval (CI) of [.855, .982] (with 10,000 bootstrap samples). Even if we included the Szmalec et al. (2011) results in this analysis (also comprising the problematic visuospatial condition), the combined correlation was .885 with a BCa 95% CI of [.710, .961] (with 10,000 bootstrap samples). Note that these numbers should be considered as lower limits as differing circumstances over studies and experiments introduce error variance. This implies that either filler performance and Hebb performance are based on the same memory processes or individual differences in the consolidation process in LTM are strongly determined by individual differences in STM processes. This finding imposes constraints on models designed to explain the relationship between immediate serial recall and the Hebb repetition effect.
Conclusions and Discussion

We begin by summarizing the main conclusion that follows from the two reported empirical studies examining a differential Hebb learning effect in dyslexic versus normal readers. Neither in the study on adolescents nor in the study on children were we able to detect a specific Hebb-learning deficit in individuals with dyslexia. Given the fact that we used a more sensitive measure to detect Hebb learning (weighted sums instead of slopes), that the groups were matched more precisely on filler performance (using state trace analysis), and given the larger power of our studies compared to Szmalec et al. (2011) (see Table 7), we can be confident that the underlying problem of dyslexics, after statistically controlling for attentional functioning, is not to be found in a specific Hebb learning problem.

As a result, we consider it more likely that it is the finding of a dyslexic Hebb learning deficit in the Szmalec et al. (2011) study that is atypical rather than our null finding. There is yet another reason for this inference. When we performed—as we did for our own data—an STA averaged out over the subjects on the Szmalec et al. data, we observed a moderate correlation (r = .44) between filler performance and Hebb performance. The lack of a strong linear relationship, however, was almost entirely due to the outlying data point of the visuospatial condition. When this data point was left out, the linear relationship increased considerably (r = .85). In the same vein, when we performed a multilevel analysis, integrating all data from the three tasks of the Szmalec study, the Bayesian factor favoring the full model (implying differential Hebb learning) was 3.45; however, when the visuospatial condition was deleted from this multilevel analysis, we observed a Bayesian factor of 2.00 favoring the null model (see Table 7). Thus, the Szmalec data of the verbal–visual and the verbal–auditory conditions together also sustained the absence of a Hebb-learning deficit in dyslexic adolescents. In conclusion, the divergent result of the Szmalec study can be pinned down to the visuospatial condition.

Second, we discuss the extent to which our results on the relationship between filler performance and Hebb performance contribute to the current discussion on the mechanisms for encoding sequence information in STM and in LTM. After effectively eliminating error variance due to individual differences by averaging out over individuals, this almost perfect linear relationship turned out to be robust over conditions (modalities) and experiments (see Figure 8). It is clear that such an intricate relationship between immediate serial recall (ISR) and long-term Hebb learning is difficult to reconcile with models conjecturing distinct and independent coding systems for both kinds of order learning. Even though (raw) Hebb performance is not the same as the Hebb effect—the latter being defined as the difference between filler and Hebb performance—the observed strength of the linear relationship leaves little room for the existence of separate coding mechanisms. From a common sense viewpoint, it can be argued that this strong relationship was largely to be expected, because the task for the participants is exactly the same when asked to recall the order of the presented filler sequence or (repeated) Hebb sequence. However, this merely begs the question: Is the improved performance on Hebb sequences the result of the processes involved in the filler performance or is it the consequence of an independent but closely related process?

Several theoretical memory models have hypothesized a distinction between a learning mechanism of ISR and an additional mechanism for long-term serial learning. In the influential model of Burgess and Hitch (2006), it is assumed that long-term serial recall is exclusively dependent on the context-timing signal, which is responsible for the encoding of the serial order of the items, whereas ISR is only mildly dependent on this signal. The main mechanism thought to be responsible for ISR is competitive queuing (Grossberg, 1987; Houghton, 1990). Hence, recovery of order information in ISR (filler sequences) is believed to be largely epiphenomenal to item coding and based on a different mechanism than recovery of order information in long-term sequence learning (Hebb sequences). Importantly, our results are not in contradiction with the idea that item information is coded differently than order information. What our results point to is that processing order information in ISR (filler sequences) has a direct influence on the longer term retention of order information (in Hebb sequences).

The connectionist model put forward by Page and Norris (2009), on which Szmalec et al. (2011) based their predictions, is probably in a better position to integrate our findings of a strong relationship between filler and Hebb performance. Although this model also implements different learning mechanisms for ISR and for long-term sequence learning, Hebb learning depends heavily on the strength of the primacy gradient encoded in the order layer (ISR), because the primacy gradient is copied into connection strengths that are controlled by a learning rate. For that reason, this model could probably be adapted to take account of individual differences in learning, in such a way that a strong relationship between filler and Hebb performance would emerge.

References

Szmalécz, A., Duyck, W., Vandierendonck, A., Mata, A. B., & Page, M. P. A. (2009). The Hebb repetition effect as a laboratory analogue of...


Received September 17, 2013

Revision received June 24, 2014

Accepted June 26, 2014