Disentangling the Contributions of Repeating Targets, Distractors, and Stimulus Positions to Practice Benefits in D2-Like Tests of Attention

Peter Wühr[1] [a], Bianca Wühr[2]

[1] TU Dortmund University, Dortmund, Germany, [2] University of Applied Sciences and Arts

When a test of attention, such as the d2 test, is repeated, performance improves. These practice benefits threaten the validity of a test because it is impossible to separate the contributions of ability and practice, respectively, to a particular result. A possible solution to this dilemma would be to determine the sources of practice effects, and to use this knowledge for constructing tests that are less prone to practice. The present study investigates the contribution of three components of a d2-like test of attention to practice benefits: targets, distractors, and stimulus configurations. In Experiment 1, we compared practice effects in a target-change condition, where targets changed between sessions, to a target-repetition condition. Similarly, in Experiment 2, we compared practice effects in a distractor-change condition to a distractor-repetition condition. Finally, in Experiment 3, we compared practice effects in a position-repetition condition, where stimulus configurations were repeated within and between tests, to a position-change condition. Results showed that repeating targets and repeating distractors contribute to practice effects, whereas repeating stimulus configurations does not. Hence, in order to reduce practice effects, one might construct tests in which target learning is prevented, for example, by using multiple targets.

Introduction

The present study investigates the sources of practice effects in pen-and-paper tests of attention, like the test d2. There are, in fact, several types or "mechanisms" of attention, such as focused (or selective) attention, divided attention, and sustained attention or vigilance (e.g., Parasuraman & Davies, 1984; Pashler, 1998). Focused or selective attention refers to the ability to find and selectively process (or respond to) relevant stimuli among irrelevant stimuli. Divided attention refers to the ability to distribute processing resources among multiple stimuli or tasks. Finally, sustained attention (or vigilance) refers to the ability to remain focused or vigilant for longer periods of time. Typical pen-and-paper tests of sustained attention, such as the d2, require visual search for relevant stimuli (called "targets") among similar, but irrelevant stimuli (called "distractors"). Because visual search requires focused attention (e.g., Treisman, 1988; Wolfe, 1998), the d2 is more a test of "focused" rather than "sustained" attention, although it is sometimes called a "sustained-attention" test (e.g., Blotenberg & Schmidt-Atzert, 2019, 2020; Steinborn et al., 2018).

The test d2 was created by Brickenkamp (1962), and is available in a revised version since 2010 (Brickenkamp et al., 2010). In its present form, the d2 consists of a sheet of paper with 14 rows of 57 stimuli each. Each stimulus consists of a combination of either the letter "d" or "p" combined with one to four dashes, which are placed above or below the letter. The three combinations of the letter "d" with two dashes form the set of target stimuli. Five combinations of the letter "d" with one, three, or four dashes, and five combinations of the letter "p" with one or two dashes form the set of distractor stimuli. The participants' task is to search each line from left to right for targets, and to mark each target with a diagonal stroke. Depending on the to-be-expected level of performance, participants are given 15 or 20 seconds per line. Two central measures of performance are computed from the raw data: the sum of hits (called "KL"), and the percentage of errors (called "F%"). The test d2 is a language-free test, and the instructions are available in different languages, including German (Brickenkamp et al., 2010), English (Brickenkamp & Zillmer, 1998), Spanish

a Correspondence concerning this article should be addressed to Peter Wühr, Technische Universität Dortmund, Institut für Psychologie, Emil-Figge-Strasse 50, 44227 Dortmund, Germany. Email: peter.wuehr@tu-dortmund.de

(e.g., Rivera et al., 2017), and Japanese (e.g., Yato et al., 2019).

Practice effects in the test d2

It has long been known that repeating the d2 leads to performance improvements in the second session (e.g., Brickenkamp, 2002; Schmidt-Atzert et al., 2004). Practice benefits in the d2 are typically in the size of 10% (improvement, as compared to the first session) in adult samples when the test is repeated after one or two weeks (e.g., J. G. Harris et al., 2007; Steinborn et al., 2018). For other pen-and-paper tests of focused attention, such as the FAIR test (Moosbrugger & Oehlschlägel, 2011), authors reported even higher practice benefits in the second session that persisted over three months (e.g., B. Wühr & Wühr, 2021). Further investigations with d2-like tests suggest that practice gains can still be observed in the seventh or eighth repetition of the test (e.g., Westhoff & Dewald, 1990).

Practice benefits have serious consequences for the validity of the test d2 (cf. Hagemeister & Westhoff, 2011; Lievens et al., 2007; Schmidt-Atzert et al., 2004). Since the investigator usually does not know the practice level of the test-taker, he or she cannot judge the contribution of practice to the test result. In other words, the investigator does not know the ratio in which ability and practice have contributed to performance, and thus cannot simply attribute performance to ability only. Another problem is that test-takers, who are expecting to be tested with the d2, can deliberately practice for the test, and thus improve their test result.

There are two ways of dealing with the problem of practice benefits in the d2. The first possibility is to develop methods for assessing the practice level of a test-taker. Unfortunately, attempts of developing such methods have not yet been successful (e.g., Hagemeister, 2007; Hagemeister et al., 2002). The second possibility is to identify the sources of practice benefits in the d2 test, and to use the results for developing new tests that are less susceptible to practice. Unfortunately, studies on the sources of practice benefits in the d2 are rare: there are only two. In a first study, Blotenberg and Schmidt-Atzert (2019) defined three component processes (perceptual speed, simple mental operation, motor speed), which are involved in doing the d2, and tried to identify which of them benefits from practice. Perceptual speed was measured in a task that required to indicate which leg of a a pi-shaped figure is longer. Motor speed was measured in a simple-reaction time task. Moreover, six computerized versions of the d2 test were constructed for assessing the "simple mental operation". The six versions of the test resulted from combining three stimulus arrangements (one stimulus, three stimuli, ten stimuli) and two paces (i.e. self-paced vs. force-paced). The authors observed practice benefits for each process, but the largest benefits were observed for the "simple mental operation". Unfortunately, the tasks used for measuring the three processes were not process-pure, because each task requires perceptual and motor speed. Moreover, it is not clear how these results can be used for constructing new tests that are less susceptible to practice. In particular,

the results of Blotenberg and Schmidt-Atzert do not show which particular component of the d2 (i.e., targets, distractors, stimulus configurations) benefits from practice. For example, their task used for measuring the "simple mental operation" always contained the same targets and distractors, and therefore the practice benefits observed for this task may have resulted from target learning, distractor learning, or both.

In another study, Wühr (2019) started to isolate components of the test d2 that benefit from practice. Therefore, he compared practice benefits in two variants of the d2 that only differed in targets. In particular, one test required searching for the letter "d" with two dashes among distractors, whereas the other test required searching for the letter "p" with two dashes among distractors. The set of distractors, which was used in both tests, contained different combinations of four letters (b, d, p, q) with different numbers of dashes. One group of participants did the same test twice with one week between sessions. Hence, for this group, every part of the test (i.e., targets, distractors, stimulus positions, motor requirements) repeated between sessions. In contrast, for another group of participants, the set of targets changed from session 1 to session 2, while all other parts of the test remained the same. Wühr observed practice benefits in both conditions, but benefits were significantly larger when all parts of the test repeated than when everything but the targets repeated. Two conclusions can be drawn from these results. First, the repetition, and thus the learning, of target processing in the first session makes a significant contribution to practice benefits in the d2. Second, learning of targets is not the only source of practice benefits in the d2, since significant practice benefits were also observed in the condition where targets changed.

The aim of the present study is to replicate and extend the findings of Wühr (2019). Therefore, we conducted three experiments in which we investigated whether repetition of different parts of the d2 improves performance. In particular, we investigated whether repetition of targets (Experiment 1), repetition of distractors (Experiment 2), and repetition of stimulus configurations (Experiment 3) improves performance in the d2. Research on visual attention and visual search have shown that the repetition of the three task components can improve performance in different tasks (e.g., Le Dantec et al., 2012). The following sections provide a summary of these results.

Attention and visual search

Pen-and-paper tests of attention, such as the d2 or FAIR-2, are visual-search tasks. These tasks have long been used for investigating visual attention (e.g., Neisser, 1963; Schneider & Shiffrin, 1977; Treisman & Gelade, 1980). In a typical visual-search task, participants have to decide if a particular, pre-defined target stimulus is present in a display of multiple stimuli or not. Hence, some displays contain a target among many distractors, whereas other displays contain only distractors. Independent variables have been the type of search task, and the number of stimuli in a display (i.e., set size; see, Chan & Hayward, 2013; Wolfe, 1998, for reviews). In feature-search tasks, the target differs

in one feature from the distractors, for example, when participants search for a red target among blue or green distractors. In contrast, in conjunction-search tasks, the target differs in a particular combination of features from the distractors, for example, when participants search for a red circle among blue circles and red squares. Typically, set size has little impact on search RTs in feature-search tasks, whereas search RTs monotonically increase with set size in conjunction-search tasks (e.g., Treisman, 1988; Treisman & Gelade, 1980). Many authors have interpreted such findings as evidence for a dichotomy between parallel search for features, which does not involve focused attention, and serial search for feature conjunctions, which requires focused attention (e.g., Treisman, 1988; Treisman & Gelade, 1980; Wolfe, 1994). In serial search, participants are assumed to direct attention onto each stimulus to process all its features, and to compare a stimulus representation with a template of the target stimuli, which is stored in memory. The d2 test requires conjunction search because the targets are defined by specific conjunctions of features, with the single features being also present in distractors. In contrast, the d2 test is not a feature-search task because a single feature does not distinguish all targets from all distractors.

Effects of repeating target stimuli

Repetition of targets in filtering tasks. When the d2 is taken twice, the target stimuli are repeated among other parts of the task. Many studies with different tasks have shown that the isolated repetition of targets can improve performance in the second trial or test (e.g., Scarborough et al., 1977). A prominent example are "positive" priming effects in so-called "filtering" tasks, which require selecting and responding to a (target) stimulus presented along with irrelevant (distractor) stimuli. In a variant of this task, participants are presented with a sequence of two trials, the prime and the probe trial. In each trial, a display containing two stimuli (e.g., letters) in different colors is presented. The green letter is the target, the red letter is the distractor. In such tasks, repeating the target while changing the distractor from prime to probe trials leads to faster responses to the probe target, as compared to a sequence in which target and distractor change (e.g., Tipper, 1985; Tipper & Cranston, 1985). If, however, the prime distractor becomes the probe target, performance is worse when compared to sequences in which target and distractor change—an observation called "negative" priming (for reviews, see, Fox, 1995; Frings et al., 2015).

Dominant accounts of priming effects in filtering tasks are dual-process theories of selective attention, and theories of episodic retrieval. Dual-process theories of selective attention assume that attentional selection involves both the amplification of targets, and the inhibition of distractor stimuli (e.g., Houghton & Tipper, 1994; Tipper, 1985). A response to the target is made when the activation of one stimulus representation (the target, in most cases) exceeds the activation of all other stimulus representations (the distractors, in most cases). After the response, the activation levels of the stimulus representations return slowly back to their resting levels. When the target is repeated before its activation has arrived at rest, target processing benefits from residual activation. When, however, the prime distractor becomes the probe target, then probe target processing will suffer from residual inhibition of the prime distractor (e.g., Frings & Wühr, 2007b; Houghton & Tipper, 1994; Tipper, 1985).

Theories of episodic retrieval assume that, after having made a response to a display, participants store stimuli and response in a memory episode of that trial (e.g., Huang et al., 2004; Neill, 1997; Neill & Valdes, 1992). The memory episode contains information about each stimulus, their role or status (as a target or distractor), and the response to each stimulus. The re-occurrence of a stimulus from a preceding trial triggers the retrieval of the most recent memory episode(s) containing this stimulus, and the content of the retrieved episode can be congruent or incongruent with the requirements of the present trial. In particular, when the target is repeated, the information stored in the memory episode of the prime trial is congruent with the requirements of the probe trial, and a response is facilitated. When however, the prime distractor becomes the probe target, the information stored in the memory episode of the prime trial is incongruent with the requirements of the probe trial, and a response is hampered (e.g., Neill, 1997; Neill & Valdes, 1992).

Priming effects observed in filtering tasks cannot be readily generalized to explain practice effects in pen-and-paper tests of attention. First, the displays in filtering tasks typically involve small numbers of stimuli, whereas tests of attention present large numbers of stimuli at once. Second, the time intervals between prime and probe trials in priming studies are very short (i.e., seconds), whereas much longer intervals (i.e., weeks, or even months) are relevant for practice benefits with attention tests. Third, the stimuli in priming experiments are usually changing their roles as targets and distractors many times during the course of an experiment (varied mapping), whereas in attention tests the stimuli maintain their roles as targets or distractors (consistent mapping). In fact, Schneider and Shiffrin (1977) have shown that varied mapping can impair learning in visual-search tasks even with many trials of practice, whereas consistent mapping can lead to massive improvements of performance in visual-search tasks (see also, Rogers & Fisk, 1991).

Repetition of targets in search tasks. With a consistent mapping of stimuli in visual search tasks, performance improves with practice (e.g., Neisser, 1963; Schneider & Shiffrin, 1977). Schneider and Shiffrin (1977) explained the improvements of search performance by search processes becoming increasingly automatic with practice. In particular, these authors assumed that extended practice with consistent mapping leads to durable changes of the attentional weights assigned to cognitive stimulus representations. According to this activation-strength model, practice produces a durable increase in the attentional weights of target representations, and a durable decrease in the attentional weights of distractor representations (see, also, Czerwinski et al., 1992; Rogers, 1992; Shiffrin, 1988). The ac-

tivation-strength model can be viewed as a variant of the dual-process model of selective attention described above (Houghton & Tipper, 1994; Tipper, 1985).

The activation-strength model can explain the finding that switching targets and distractors after extended practice leads to a massive decrement in performance (e.g., Prinz, 1979; Rogers, 1992; Rogers & Fisk, 1991; B. Wühr & Wühr, 2021). The explanation is that the attentional weights of targets and distractors must be slowly re-learned after having been switched. Additional studies have investigated how changing only the targets in a visual-search task after extended practice affects performance. In one study, Fisk, Lee, and Rogers (1991) used a task in which participants searched for a target word (defined by category) among a display of three words from different categories. Participants were trained for ten sessions with a consistent mapping (e.g., with target set A and distractor set B). After practice, the authors compared performance in three transfer conditions: A target-repetition condition (target set A, distractor set C), a target-becomes-distractor condition (target set C, distractor set A), and a control condition (target set C, distractor set D). Results showed that performance in the target-repetition condition was (significantly) better than in the control condition, whereas performance in the target-reversal condition was (numerically) worse than in the control condition. The pattern is consistent with the attention-strength model, and demonstrates that learning target stimuli makes an important contribution to practice effects in visual search (see, also, Le Dantec et al., 2012; Rogers, 1992; P. Wühr, 2019).

The beneficial effects of repeating targets in laboratory search tasks, and in pen-and-paper tests of attention (e.g., Fisk et al., 1991; Le Dantec et al., 2012; Rogers, 1992; P. Wühr, 2019) can also be explained by episodic retrieval (e.g., Logan, 1988, 1990; Neill, 1997; Neill & Valdes, 1992). For example, one might assume that participants store every processing episode in a continuous pen-and-paper test in memory. With practice under consistent-mapping conditions, participants will store increasing numbers of previous target and distractor episodes. When encountering a new target or distractor, congruent episodes are retrieved from memory and facilitate processing of the current stimulus.

Effects of repeating distractor stimuli

Repetition of distractors in filtering tasks. When the d2 is taken twice, the distractor stimuli are repeated among other parts of the task. Results of studies with filtering tasks have shown that the repetition of distractors from prime to probe displays improve performance, independently from whether the target is also repeated or not (e.g., Frings & Wühr, 2007a; Neumann & DeSchepper, 1991; Tipper et al., 1989; Tipper & Cranston, 1985). Both the dual-process theory of attention and episodical-retrieval theory can explain the beneficial effects of distractor repetition. According to the former, processing of the probe target benefits from lingering inhibition of the repeated distractor. According to the latter, the repetition of the distractor stimulus triggers the retrieval of a previous processing episode (prime trial)

that is congruent with the requirements of the probe trial (e.g., Frings & Wühr, 2007a). Yet, the effects of distractor repetition in filtering tasks have been shown in non-search tasks with very short intervals between prime and probe displays. Hence, it is unclear whether these findings generalize to the repetition of visual search tasks with long time intervals between sessions.

Repetition of distractors in search tasks. Studies of practice effects in visual-search tasks, with a consistent mapping, have shown that changing the distractors from practice to test deteriorates performance (e.g., Fisk et al., 1991; Le Dantec et al., 2012; Rogers, 1992). In the study by Fisk et al. (1991) participants practiced for ten days with a target set A, and a distractor set B. On day 11, participants were tested in three different transfer conditions: the distractor-repetition condition (target set C, distractor set B), a distractor-becomes-target condition (target set B, distractor set C), and a control condition with untrained stimulus sets. Results showed that performance in the distractor-repetition condition was (significantly) better than in the control condition, whereas performance in the distractor-becomes-target condition was (numerically) worse than in the control condition (cf., Rogers, 1992, for similar results). This pattern of findings implies that learning distractor stimuli also contributes to practice effects in visual search (see, also, Le Dantec et al., 2012). The beneficial effects of repeating distractors in visual search are consistent both with the activation-strength model (Rogers, 1992; Schneider & Shiffrin, 1977; Shiffrin, 1988), and with episodic-retrieval theories (e.g., Logan, 1988, 1990; Neill, 1997).

Effects of repeating stimulus positions

Repetition of stimulus positions in filtering tasks. The fourteen stimulus lines in the d2 are repetitions of lines 1-3. That is, line 1 is identical with lines 4, 7, 10 and 13. Similarly, line 2 is identical with lines 5, 8, 11, and 14, and line 3 is identical with lines 6, 9, and 12. The repetition of lines means that particular spatial configurations of targets and distractors are repeated in a session with the d2, and these configurations could thus be learned. In fact, findings from studies with filtering tasks show that repeating combinations of targets and locations from prime to probe trials improves performance in the probe trial, as compared to conditions in which either the target or the location changes (e.g., Chao & Yeh, 2005; Guy & Buckolz, 2007; Park & Kanwisher, 1994).

Repetition of stimulus positions in search tasks. Research on "contextual cueing" has shown that people can implicitly learn complex stimulus configurations in visual search tasks. In an influential study, Chun and Jiang (1998; see, also, Chun, 2000) showed that repeating spatial configurations of stimuli improves search performance as compared to non-repeated configurations. In their Experiment 1, participants searched for the rotated target letter "T" among heterogeneously rotated distractor letters "L". Each display contained a target among distractors, and participants reported the orientation of the target by pressing a key. Each block of trials contained 12 trials with "old" dis-

plays, which were repeated throughout the experiment, and 12 trials with "new" displays, which were not repeated. The total of 30 experimental blocks were divided in six epochs of 5 blocks each. Results showed that, starting with epoch 2, RTs were shorter to old as compared to new displays. This "contextual cueing" effect is explained by assuming that participants quickly learn the position of the target in repeated (i.e., old) displays, which facilitates search and a response to the target in subsequent repetitions of the display. Further experiments suggested that contextual cueing involves learning of spatial configurations, whereas participants do not seem to learn much about distractor identities (e.g., Chun & Jiang, 1998; but see Makovski, 2016, for diverging results). Later studies on the mechanisms of the contextual cueing effect suggest that contextual cueing facilitates attentional control of visual search, and not the response to the target (e.g., A. M. Harris & Remington, 2017; see, Sisk et al., 2019, for a review).

Independent effects of repeating targets, distractors, and positions

The studies reviewed in the previous sections have independently shown that the repetition of targets, distractors, or stimulus positions between two trials or tests can facilitate visual search, as compared to changing these components. Hence, the question arises as to whether repeating different components of a search task has independent effects, or whether these effects may interact with each other. Le Dantec et al. (2012) addressed this question in an interesting study, in which participants practised a visual search task for ten days. Targets and distractors were lines differing in orientation. During practice, participants searched for a target with a particular orientation (either 45° or 135°) among heterogeneous distractors with similar orientations. To address contextual learning, the practice sessions included the same amount of "old" (i.e., repeated) and "new" (i.e., unrepeated) configurations. During practice, participants showed both effects of stimulus (i.e., target and distractor) learning, and effects of contextual learning. Stimulus learning was reflected in the fact that search times continuously decreased, and search accuracy increased, during learning. Contextual learning was reflected in the fact that search performance was better for repeated as compared to unrepeated displays. On day 11, participants were finally tested with practiced and unpracticed targets, with practiced and unpracticed distractors, and with practiced and unpracticed stimulus configurations. Results showed that practice benefits were confined to trained stimuli and trained configurations. In particular, better test performance when searching for practiced as compared to unpracticed targets revealed target learning. Better test performance when searching among practiced as compared to unpracticed distractors revealed distractor learning. Moreover, and interestingly, the authors did not observe any interactions between target learning, distractor learning, and contextual learning. Hence, targets, distractors, and stimulus configurations were learned independently, which may suggest that different mechanisms subserve these learning and practice effects (see, also, Geng et al.,

2019). The results of Le Dantec et al. (2012) are important for our study because they justify our approach of independently investigating the effects of learning targets, distractors, and spatial configurations.

The present study

The aim of the present study is to find out which components of a pen-and-paper test of attention are learned in a single testing session, and may therefore improve performance in a second testing session when the component is repeated. Therefore, we conducted three experiments with different custom-made variants of the d2 test. In Experiment 1 we addressed the repetition of targets, in Experiment 2 we addressed the repetition of distractors, and in Experiment 3 we finally addressed the repetition of stimulus configurations. Basic research on visual attention and visual search has shown that the repetition of targets (e.g., Neill, 1977; Rogers, 1992), the repetition of distractors (e.g., Frings & Wühr, 2007a; Rogers, 1992), and the repetition of stimulus positions and configurations (e.g., Chun & Jiang, 1998; Park & Kanwisher, 1994) can improve performance in search and in non-search tasks. With the exception of targets (P. Wühr, 2019), it is not clear whether the findings from basic research generalize to repetitions of the d2 and similar tests, because of many methodological differences between the tasks. The most obvious differences concern time intervals between practice and test, display sizes, and test formats. Further differences relate to whether the search is self-paced or not, and whether participants can preview subsequent items or not (e.g., Blotenberg & Schmidt-Atzert, 2020).

The results of the present experiments might reveal useful insights for the construction of focused-attention tests that are less susceptible to practice effects than existing tests. If, for example, we observe that the repetition of stimulus rows in the d2 leads to superior performance than a test without repetitions of stimulus rows, then using the latter type of test would help to reduce practice effects.

We would also like to stress that testing between different explanations for practice effects in visual-search tasks is *not* an aim of this study. We did not design the present experiments for testing between theories of practice effects or automatization in visual search. This does not mean, however, that the results of the present experiments will be completely mute about this issue. Therefore, we will discuss some implications of our findings for evaluating theories of practice in visual search.

Experiment 1

In Experiment 1, we investigated whether repeating targets between two sessions contributes to practice effects in a d2-like test. Therefore, we tested two groups of participants two times with custom-made variants of the d2 with one week between sessions. The first group did the same version of the test in both sessions (target-repetition condition). Hence, in this condition, all relevant components of the test (targets, distractors, stimulus positions and configurations) repeated from session 1 to session 2. For the

second group, the set of targets changed from session 1 to session 2, whereas all other components of the test (i.e., distractors, stimulus positions and configurations) were repeated.

If participants selectively practice detecting and processing targets in the first session with the d2, the results of practice should improve performance in the second session with the d2 when the targets repeat, but not when the targets change (Le Dantec et al., 2012; P. Wühr, 2019). Although we predicted larger practice effects in the target-repetition condition than in the target-change condition, we also expected significant practice benefits in the target-change condition. Yet, several sources might be responsible for practice benefits in the target-change condition, such as the repetition of distractors, the repetition of stimulus positions and configurations, as well as stimulus-unrelated test features.

Experiment 1 was an attempt to replicate and extend the findings reported by Wühr (2019). We considered this replication important for three reasons. First, we wanted to investigate the contribution of the most important components of the d2 test (targets, distractors, stimulus positions) in the same series of experiments. Second, the sample in the experiment by Wühr (2019) was small, and we aimed at replicating his findings with more power. Third, we wanted to have the possibility of comparing practice benefits for three most important components of the d2 test, and therefore the experiments should have both sufficient and similar power.

Method

Participants. For this and the subsequent studies, we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. In a previous study with 48 participants, the critical two-way interaction had an effect size of partial eta² = .127 (P. Wühr, 2019). A power analysis with a tool provided by the website "StatistikGuru" (Hemmerich, 2015–2022) revealed that a sample size of 122 (i.e., 61 per group) would be required to obtain the desired effect with high power (1-β = .95, when α = .05) in a two-factorial mixed Analysis of Variance (ANOVA). A total of 159 volunteers (135 women, 24 men; mean age = 21.88 years, SD = 4.20) participated in a first session, and 142 participants completed two sessions. We obtained informed consent from each participant before inclusion in the study; participants received course credits for their time. The two tests (A, B) were equally distrib-

uted between participants, and participants were randomly assigned to the two conditions. All procedures used in the present experiments were consistent with the 1964 Helsinki declaration and its later amendments, and with the Ethical Research Guidelines by the German Society of Psychology (Deutsche Gesellschaft für Psychologie).[1]

Materials. We constructed two pen-and-paper tests that closely resembled the d2. Each test consisted of 16 rows with 54 stimuli. While the same set of distractors was used for all tests, the set of targets differed between the two tests. In particular, for test A, the target set included a "d" with two dashes below, a "d" with two dashes above, and a "d" with one dash below and one above (i.e., the same set of targets as in the original test). For test B, the target set included a "p" with two dashes below, a "p" with two dashes above, and a "p" with one dash above and one below. The distractor set always included a "d" with one dash above, a "d" with one dash below, and a "d" with two dashes above and two below. In addition, the distractor set also included the letters "b", "p" and "q" with two dashes above, or two dashes below, or one dash above and one below, respectively. Hence, each stimulus set included three targets and twelve distractors. In each row, each target was presented six times and each distractor was presented three times in random order. Hence, each row included 18 targets, and 36 distractors. Excluding the first line, which was considered for practice, the complete test consisted of 810 stimuli with 270 targets, and 540 distractors.[2] We did not conduct a pilot study for Experiment 1 because similar tasks had already been used in Wühr (2019), and ceiling effects were very rare in that study.

The stimuli were printed in font Helvetica with a font size of 10. The horizontal distance between two stimuli was 0.5 cm, the vertical distance between two stimuli was 1.0 cm. At the right side of the test page, a table with 16 rows and three columns for filling in counts of hits and errors was printed. The tests were printed on the backside of a sheet of paper with size DinA4. The upper half of the frontside contained a box with several lines for filling in demographic information (i.e., name, date of birth, gender, and highest degree). In the lower half, a short instruction was presented together with a depiction of the three target stimuli for this condition. Then there was a row of 30 stimuli (i.e., 3 targets and 12 distractors, each presented twice) for practice. Below the practice line were three empty boxes for filling in counts of hits and errors.

---

1 When this study was planned in 2018, the ethical guidelines of the German Psychological Society (Deutsche Gesellschaft für Psychologie, DGPs) did not require ethical approval for studies in which only behavioral-performance data were collected. Therefore, we did not seek ethical approval for the experiments reported in this study.

2 The number of distractors and the target-to-distractor ratio in our version of the test were different from the original test d2-R. The original test contains 10 different distractors that occur with different frequencies, a fact that is not explained in the manual. Overall, without line 1 and 14, the original test consists of 308 targets and 376 distractors. Hence, the target-to-distractor ratio is 1:1.2. In contrast, the target-to-distractor ratio was 1:2 in Experiment 1, and 1:1.7 in Experiment 2. The target-to-distractor ratio does not seem to affect practice benefits in the d2 test of attention because we observed very similar practice benefits in the hit rate (KL) observed in the complete-repetition conditions of the three experiments despite different target-to-distractor ratios.

Procedure. The participants were tested in large groups in lecture halls. Each participant was tested twice with one week between the two sessions. At the beginning of the first session, all participants provided informed consent to participate in the study. After the test sheets had been handed out, instructions were read to the participants. The instructions closely followed those for the d2-R (Brickenkamp et al., 2010). In particular, participants were told to mark a target with a single stroke, to work as quickly as possible, and to avoid errors, too. Notably, however, in contrast to the original instructions, participants were told not to correct errors because correcting errors might disrupt search. Then participants practiced their task on a row of 30 stimuli. After practice, participants were instructed to turn the page and to start working on the first line of the test. After 15 seconds, the experimenter said "Stop. Next line", and participants switched to the beginning of the next row. This procedure was repeated until participants had finished the test. The procedure for the second session was similar to the first, except that (all) participants were told that some of them might have to search for the same targets as in the first session, whereas others might have to search for new targets.

For the first session, the two versions of the test (differing in target sets) were randomly distributed among the participants. Each test had a unique participant number that determined the testing condition. One half of the participants was assigned to the target-repetition condition, whereas the other half was assigned to the target-change condition. Participants in the target-repetition condition received two tests with the same target set (i.e., A-A, B-B), whereas participants in the target-change condition received two tests with different target sets (i.e., A-B, B-A).

Design and data analysis. The data analysis was preregistered at OSF after testing was completed, but before data analysis was begun (osf.io/w6qej). The analyses were based on two-factorial mixed designs with Session (first session, second session) as a within-subjects factor, and Condition (target repetition, target change) as a between subjects-factor. We analyzed the impact of these factors on three dependent variables: number of hits, percentage of false alarms, and percentage of misses. For computing error percentages, we used the number of inspected items (called "GZ") per test in the denominator. The number of inspected items is the sum of all items to the left of the last marked stimulus in a line. If the two-way interaction was significant, we planned to conduct pair-wise comparisons for determining the source of the interaction. We used one-tailed tests for pair-wise comparisons because we had specific hypotheses for these comparisons. In particular, we always expected that the repetition condition would produce stronger practice benefits from the first to the second session than the change condition.

Note that our error analyses deviated in two aspects from the usual error analyses of the d2 test. Firstly, we separately analyzed false alarms and misses, whereas both measures are collapsed into a single error-percentage measure in the usual analysis. The main reason was that false alarms reflect the attentional strengths of distractor items,

whereas misses reflect the attentional strengths of targets, and therefore their separate analysis may reveal additional information about the processing of targets and distractors. Secondly, we used GZ in the denominator for computing error percentages, whereas the number of hits (plus errors) was used in the denominator for the usual analysis. Our main reason is that GZ includes both targets and distractors, and therefore provides a more adequate baseline for computing error percentages than hits, which include only targets.

Before analyzing the data, we eliminated incomplete data sets and data sets with outliers in dependent variables. Data sets were incomplete when (a) participants only performed the first session, or (b) skipped lines in the test. Moreover, we checked the data sets of the *first* session for outliers on either the number of hits or the percentage of errors (collapsed across false alarms and misses). We excluded data sets that violated the Tukey criterion on one of the two variables. In particular, Tukey (1977) defined an outlier as a value that lies outside of [Q1 – 1.5 × IQR; Q3 + 1.5 × IQR]. We will report the frequency of these cases at the beginning of the results section for each experiment.

Results

Data exclusion and test reliability. After data collection was completed, we excluded two participants because their tests were not complete (i.e., they skipped lines), and seven additional participants because their performance violated the Tukey outlier criterion. Three of the excluded participants achieved less than 50 hits, and five made more than 80 errors in the first session. Hence, there were 133 participants in the final sample (113 women, 20 men; mean age = 21.865), with 68 participants in the target-repetition condition and 65 participants in the target-change condition. Test-retest reliabilities of test scores, and correlations between different test-scores within sessions, are shown in table A1 in the Appendix.

Hit rate. **Table 1** shows summary statistics for several performance scores as a function of Session (1, 2) and Condition (target repetition, target change). The condition means of the hit rates are shown in **Figure 1A**. Shapiro Wilk tests showed that the distribution of hit rates did not significantly differ from normal in either session, both $W > 0.970$, both $p > .40$. The participants' hit rates were submitted to a mixed-model ANOVA with *Session* as within-subjects factor and *Condition* as between-subjects factor. The main effect of Session was significant, $F_{(1, 131)} = 112.065$, $MSE = 106.338$, $p < .001$, $\eta_p^2 = .461$, indicating higher hit rates in session 2 than in session 1. The main effect of Condition was also significant, $F_{(1, 131)} = 4.076$, $MSE = 653.265$, $p = .046$, $\eta_p^2 = .030$. Most importantly, the Session × Condition interaction was also significant, $F_{(1, 131)} = 20.401$, $MSE = 106.338$, $p < .001$, $\eta_p^2 = .135$. The interaction reflected the finding that participants in the target-repetition condition showed a larger increase in hit rate from session 1 to session 2 (*Mean* difference = 19.103, $SD = 14.633$; $t[67] = 10.765$, $p < .001$, $d = 1.305$) than participants in the tar-

Table 1. Summary statistics for performance scores observed in Experiment 1 as a function of Session (1, 2), and condition (target repetition, target change).

	Hits		False Alarms		Misses	
	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2
Target-repetition	115.309 (28.437)	134.412 (33.249)	0.555	0.269	5.482	4.362
Target-change	114.692 (28.476)	122.369 (30.413)	0.715	0.359	5.441	4.420
Total	115.001 (28.456)	128.391 (31.863)	0.635	0.314	5.462	4.391

Hits are given in absolute numbers, and as percentages (in brackets) in relation to the total number of inspected items in session 1. False alarms and misses are given as percentages in relation to the total number of inspected items in the corresponding session.

get-change condition (*Mean* difference = 7.952, *SD* = 14.679; $t[64] = 4.259$, $p < .001$, $d = 0.528$).

Error scores. The percentages of false alarms and the percentages of misses were separately analyzed. Shapiro Wilk tests showed that the distribution of all error percentages deviated significantly from normal, all Ws < .95, all ps < .020. Therefore, we analyzed error percentages with non-parametric tests. Wilcoxon tests showed that both false alarms, $W = 4914.000$, $p < .001$, $r_b = .499$, and misses, $W = 6409.000$, $p < .001$, $r_b = .438$, decreased from session 1 to session 2. The decrease in false alarms was numerically smaller for the target-repetition group than for the target-change group, but this difference was not significant, Mann-Whitney's $U = 2088.500$, $p = .709$, $r_b = .055$. The decrease in misses was numerically larger for the target-repetition group than for the target-change group, but this difference was not significant, too, Mann-Whitney's $U = 2067.000$, $p = .741$, $r_b = .065$.

Discussion

Experiment 1 investigated whether repeating targets between two sessions contributes to practice effects in a d2-like test. Therefore, we compared changes in performance between two sessions in a target-repetition condition and a target-change condition. In both conditions, we observed significant improvements from session 1 to session 2 in all measures of performance. In addition, we observed numerically larger improvements in the target-repetition condition, as compared to the target-change condition, in hits (KL) and in misses, with the difference being significant for hits. In fact, for hits, performance improved more than twice as much in the target-repetition condition (improvement = 16.6%) than in the target-change condition (improvement = 6.7%). Hence, we successfully replicated the results of Wühr (2019), who reported a similar difference in improvements (i.e., 15.4% in the target-repetition condition versus 7.4% in the target-change condition). In summary, results of Experiment 1 suggest that participants selectively learn and improve the processing of targets in session 1, and this learning improves performance in a subsequent session with the same targets, but not in a subsequent session with different targets.

Experiment 2

In Experiment 2, we investigated whether repeating distractors between two sessions contributes to practice effects in a d2-like test. As in Experiment 1, we tested two groups of participants two times with custom-made variants of the d2 with one week between sessions. The first group did the same version of the test in both sessions (distractor-repetition condition). Hence, in this condition, all relevant components of the test (targets, distractors, stimulus positions and configurations) repeated from session 1 to session 2. For the second group, the set of distractors changed from session 1 to session 2, whereas all other components of the test (i.e., targets, stimulus positions and configurations) were repeated.

If participants can also learn rejecting distractors in the first session with the d2, the results of this practice should improve performance in the second session with the d2 when the distractors repeat, but not when the distractors change (Le Dantec et al., 2012; Rogers, 1992). In fact, the results from laboratory studies suggest that repeating distractors between two trials or sessions can improve performance, as compared to changing distractors, both in non-search tasks (e.g., Frings & Wühr, 2007a; Tipper & Cranston, 1985), and in search tasks (Geng et al., 2019; Rogers, 1992). Particularly the results of experiments with extended practice in search tasks suggest that, during practice, participants improve their ability to reject distractors (e.g., Geng et al., 2019; Le Dantec et al., 2012; Rogers, 1992; Schneider & Shiffrin, 1977). It is, however, unclear whether the results obtained in laboratory studies would also occur in pen-and-paper tests because there are a lot of methodological differences, as described above.

Method

Participants. We aimed for a comparable sample size as in Experiment 1 (i.e., 122). A total of 122 participants completed a first session, and 106 participants completed two sessions within one week. This final sample encompassed 90 female and 16 male students of different majors (mostly psychology or education). The average age of the sample was 20.58 years (SD = 2.54). Participants who completed
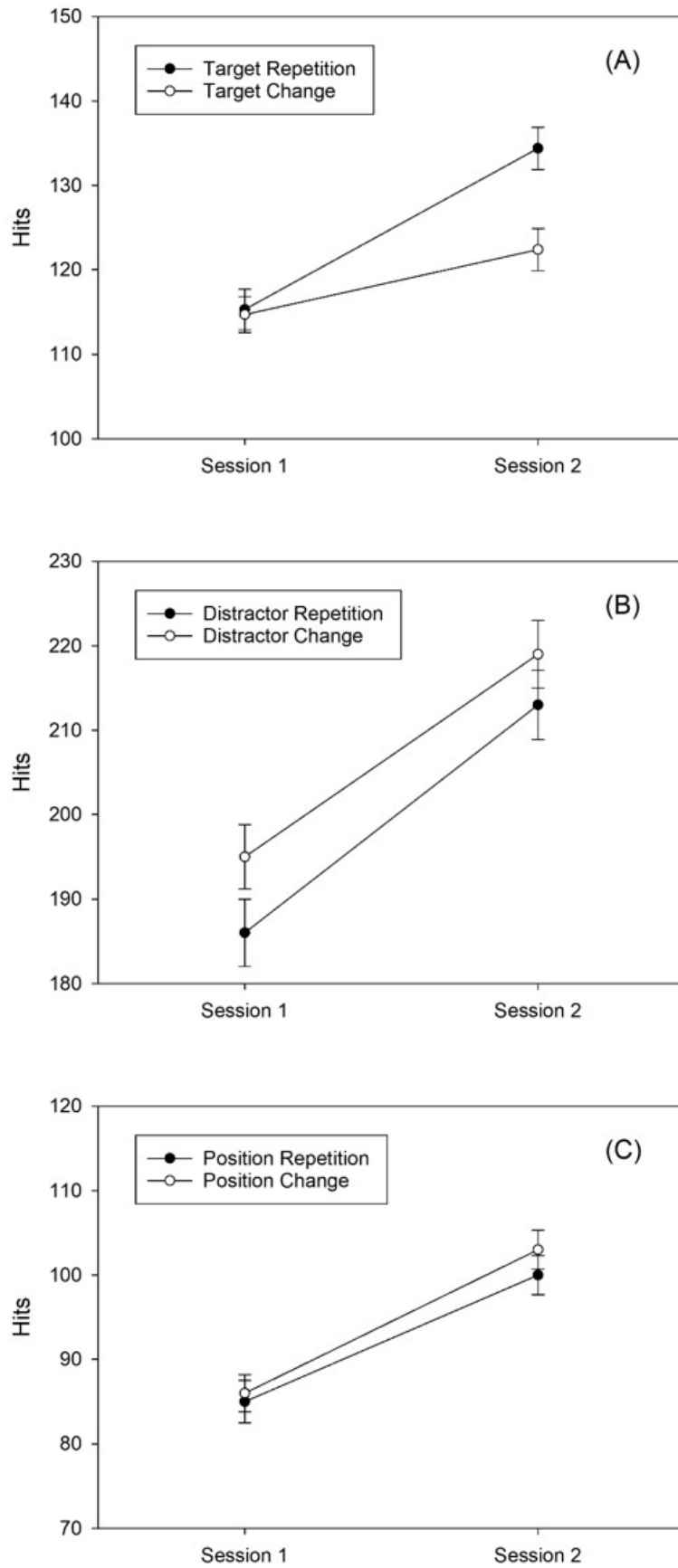
Figure 1. Hit rates observed in Experiment 1 (panel A), Experiment 2 (panel B), and Experiment 3 (panel C), as a function of Session and Condition.

Error bars represent standard errors between participants.

two sessions were compensated with course credits or coffee vouchers.

Materials. We constructed two new pen-and-paper tests that resembled the d2. Each test consisted of 16 rows with 56 stimuli. The stimulus set for each test consisted of three targets and five distractors. While the same set of targets (i.e., a d with two dashes) was used for all tests, the set of distractors differed between the two tests. In particular, for test A, the distractor set included a "d" with a dash below, a "d" with two dashes below and one above, and three "p"s with two dashes. In contrast, for test B, the distractor set included a "d" with a dash above, a "d" with two dashes above and one below, and three "q"s with two dashes. In each row of the test, each stimulus was presented seven times in a random order. Hence, each row included 21 targets, and 35 distractors. Excluding the first line, which was considered for practice, the complete test consisted of 840 stimuli with 315 targets, and 525 distractors.

There were two versions of each test (i.e., A1, A2, B1, B2). The two versions only differed in the order of stimuli in each line. At the right side of the test page, a table with 16 rows and three columns for filling in counts of hits and errors was printed. The tests were printed on the backside of a sheet of paper with size DinA4. The frontside was similar to that of the test used for Experiment 1.

Procedure. We conducted a pilot study with 12 participants to determine the probability of ceiling effects when participants were given 20 seconds per stimulus row. In this pilot study, the average number of processed stimuli per row was 36 (SD = 6). In only 2% of the cases, participants processed more than 50 stimuli in a row within 20 seconds. From these findings, we concluded that providing 20 seconds per row would not produce ceiling effects in the first session and, thus, leave enough space for observing practice benefits in the second session.

For the first session, the four versions of the test (i.e., A1, A2, B1 and B2) were randomly distributed among the participants. Each test had a unique participant number that determined the testing condition. One half of the participants was assigned to the distractor-repetition condition, whereas the other half was assigned to the distractor-alternation condition. Participants in the distractor-repetition condition received two tests with the same distractor set, but with a different spatial arrangement of stimuli (e.g., A1-A2 or A2-A1). In contrast, participants in the distractor-alternation condition received two tests with different distractor sets and with different spatial arrangements of stimuli (e.g., A1-B1 or B2-A1). Participants were not informed about this manipulation. All other features of the procedure were the same as in Experiment 1. We changed the spatial arrangement of stimuli between sessions as an additional means for preventing ceiling effects in session 2, since we did not know the results of Experiment 3 when conducting Experiment 2.

Design and data analysis. The data analysis was preregistered at OSF after testing was completed, but before data analysis was begun (osf.io/w7rua). The analyses were based on two-factorial mixed designs with Session (first session, second session) as a within-subjects factor, and

Condition (distractor repetition, distractor change) as a between subjects-factor. All other aspects of the design and data analyses followed the scheme described for Experiment 1.

Results

Data exclusion and test reliability. After data collection was completed, we excluded four participants because their performance violated the Tukey outlier criterion. Three of the excluded participants made more than 90 errors in the first session. Hence, there were 102 participants in the final sample (89 women, 13 men; mean age = 20.44), with 51 participants in the distractor-repetition condition and 51 participants in the distractor-change condition. Test-retest reliabilities of test scores, and correlations between different test-scores within sessions, are shown in table A2 in the Appendix.

Hit rate. **Table 2** shows summary statistics for several performance scores as a function of Session (1, 2) and Condition (distractor repetition, distractor change). The condition means of the hit rates are also shown in **Figure 1B**. Shapiro Wilk tests showed that the distribution of hit rates did not significantly differ from normal in either session, all $W > 0.965$, all $p > .180$. The participants' hit rates were submitted to a mixed-model ANOVA with *Session* as within-subjects factor and *Condition* as between-subjects factor. The main effect of Session was significant, $F(1, 100) = 297.338$, $MSE = 110.847$, $p < .001$, $\eta_p^2 = .748$, indicating higher hit rates in session 2 than in session 1. The main effect of Condition, $F(1, 100) = 2.149$, $MSE = 1493.223$, $p = .146$, $\eta_p^2 = .021$, and the Session × Condition interaction, $F(1, 100) = 0.854$, $MSE = 110.847$, $p = .358$, $\eta_p^2 = .008$, were not significant.

Error scores. The percentages of false alarms and the percentages of misses were separately analyzed. Shapiro Wilk tests showed that the distribution of all error percentages deviated significantly from normal, all $Ws < .94$, all $ps < .010$. Therefore, we analyzed error percentages with nonparametric tests. Wilcoxon tests showed that both false alarms, $W = 602.000$, $p < .001$, $r_b = .625$, and misses, $W = 1146.000$, $p < .001$, $r_b = .729$, decreased from session 1 to session 2. The decrease in false alarms was significantly larger for the target-repetition group (mean decrease = 0.188) than for the target-change group (men decrease = 0.040), Mann-Whitney's $U = 1014.000$, $p = .025$, $r_b = .220$. The decrease in misses was numerically larger for the target-repetition group (mean decrease = 1.519) than for the target-change group (mean decrease = 0.986), Mann-Whitney's $U = 1071.000$, $p = .063$, $r_b = .176$.

Discussion

Experiment 2 investigated whether repeating distractors between two sessions contributes to practice effects in a d2-like test. Therefore, we compared changes in performance between two sessions in a distractor-repetition condition and a distractor-change condition. In both conditions, we observed significant improvements from session 1 to session 2 in all measures of performance. In addition,

Table 2. Summary statistics for performance scores observed in Experiment 2 as a function of Session (1, 2), and condition (distractor repetition, distractor change).

	Hits		False Alarms		Misses	
	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2
Distractor-repetition	185.863 (33.779)	212.647 (38.758)	0.330	0.143	5.295	3.776
Distractor-change	195.157 (34.640)	219.216 (39.022)	0.212	0.172	4.566	3.580
Total	190.510 (34.210)	215.932 (38.890)	0.271	0.158	4.926	3.678

Hits are given in absolute numbers, and also as percentages (in brackets) in relation to the total number of inspected items in session 1. False alarms and misses are given as percentages in relation to the total number of inspected items in the corresponding session.

we observed numerically larger improvements in the distractor-repetition condition, as compared to the distractor-change condition, in all measures, with the difference being significant for false-alarm rates. In fact, the false alarms decreased more strongly from session 1 to session 2 in the distractor-repetition condition than in the distractor-change condition. The fact that the observed differences between the two conditions was numerically small is not too surprising because false alarms are relatively rare events in the d2. Observing an impact of distractor learning in false alarms, and not in misses, makes sense because false alarms reflect an error in rejecting distractors, whereas misses reflect an error in detecting targets.

The effects of distractor practice observed in Experiment 2 were smaller than the effects of distractor learning observed in Experiment 1. We see two possible reasons for the differences. First, the number of to-be-learned targets (i.e., three) in Experiment 1 was smaller than the number of to-be-learned distractors (i.e., five) in Experiment 2. Second, it is also possible that the two target sets in Experiment 1 were perceptually more distinct than the two distractor sets in Experiment 2. In fact, in Experiment 1, targets were changed by replacing the letter "d" by the letter "p" (or vice versa), whereas, in Experiment 2, distractors were changed by replacing the letter "p" by the letter "q" (or vice versa). There is evidence that the letters "d" and "p" are less similar, or more distinct, than the letters "p" and "q" (e.g., Boles & Clifford, 1989), and this difference in distinctivness might have affected the rate of transfer, and thus, practice benefits in the two change conditions. Despite these differences, the results of Experiment 2 suggest that participants can learn and improve the rejection of distractors in session 1, and this learning improves performance in a subsequent session with the same distractors, but not in a subsequent session with different distractors.

Experiment 3

In Experiment 3, we investigated whether repeating stimulus configurations within and between tests contributes to practice effects in a d2-like test. As in the preceding experiments, we tested two groups of participants two times with custom-made versions of the d2 with one

week between sessions. In both versions of the test, lines 1-3 were repeated several times until all lines were filled, as in the original d2 test, but with different stimulus configurations in the two versions of the test. The first group did the same version of the test in both sessions (position-repetition condition). Hence, in this condition, all relevant components of the test (targets, distractors, stimulus configurations) repeated from session 1 to session 2. The second group did different versions of the test in the two sessions (position-change condition). Hence, in the latter condition, stimulus configurations changed between sessions, but stimuli (i.e., targets and distractors) remained the same.

If participants can learn the positions of stimuli (targets and distractors) from several repetitions of the same rows in session 1, the results of this practice should improve performance in the second session when the stimulus configurations are repeated, but not when the stimulus configurations change. In fact, the results from laboratory studies suggest that repeating stimulus configurations between two trials or sessions can improve performance, as compared to unrepeated configurations, both in non-search tasks (e.g., Chao & Yeh, 2005; Park & Kanwisher, 1994), and in search tasks (Chun & Jiang, 1998; Le Dantec et al., 2012). Notably, effects of contextual learning have been observed after only five repetitions of the "old" stimulus configurations (e.g., Chun & Jiang, 1998; Le Dantec et al., 2012). But, as it was the case for the preceding experiments, it is yet unclear whether the results obtained in laboratory studies would also occur in pen-and-paper tests because of many methodological differences, including shorter intervals between trials or sessions, and smaller set sizes in laboratory studies.

Method

Participants. We aimed for a similar sample size as in Experiment 1. A total of 151 participants completed a first session, and 116 participants completed two sessions within one week. This final sample encompassed 87 female and 29 male students of different majors (mostly psychology or education). The average age of the sample was 22.39 years (SD = 4.56). Participants who completed two sessions were compensated with course credits or coffee vouchers.

Materials. We constructed two new pen-and-paper tests that resembled the d2. Each test consisted of 16 rows with 52 stimuli. The stimulus set for each test consisted of three target items and ten distractor items. The three targets were a "d" with two dashes above, a "d" with two dashes below, and a "d" with one dash above and one below. The distractors were "d"s with one dash (above or below), "d"s with three dashes (one above and two below, or vice versa), a "d" with four dashes (two above and two below), "p"s with one dash (above or below), and "p"s with two dashes (two below, two above, one below and one above). In each row of the test, each stimulus item was presented four times. Hence, each row included 12 targets and 40 distractors. Excluding the first line, which was considered for practice, the complete test consisted of 780 stimuli (180 targets, 600 distractors).

For test version A, we generated three rows, in which stimuli were randomly ordered, and these rows were subsequently repeated until 16 rows were created. For test version B, we generated three lines with different order of items as in test A, and again repeated these lines until 16 rows were created. Hence, in both test versions, row 1 was repeated six times, whereas rows 2 and 3 were repeated five times. The reason for the difference was that five repetitions of each row should enter the analysis. All other aspects of the tests were similar to the tests used in Experiment 1.

Procedure. For the first session, the two versions of the test (i.e., A and B) were randomly distributed among the participants. Each test had a unique participant number that determined the testing condition. One half of the participants was assigned to the position-repetition condition, whereas the other half was assigned to the position-change condition. Participants in the position-repetition condition received the same test twice (i.e., A-A, B-B), whereas participants in the position-change condition received different versions of the test in different sessions (i.e., A-B, B-A). Participants were not informed about this manipulation. All other features of the procedure were the same as in Experiment 1.

Design and data analysis. The data analysis was preregistered at OSF after testing was completed, but before data analysis was begun (osf.io/gd9y8). The analyses were based on two-factorial mixed designs with Session (first session, second session) as a within-subjects factor, and Condition (position repetition, position change) as a between subjects-factor. All other aspects of the design and data analyses followed the scheme described for Experiment 1.

Results

Data exclusion and test reliability. After data collection was completed, we excluded one participant, who had marked p's instead of d's, and four additional participants because their performance violated the Tukey outlier criterion. One excluded participant achieved less than 50 hits, and three other made more than 40 errors in session 1. Hence, there were 111 participants in the final sample (83 women, 28 men; mean age = 22.33), with 51 participants in the position-repetition condition and 60 participants in the position-change condition. Test-retest reliabilities of test scores, and correlations between different test-scores within sessions, are shown in table A3 in the Appendix

Hit rate. **Table 3** shows summary statistics for several performance scores as a function of Session (1, 2) and Condition (position repetition, position change). The condition means of the hit rates are also shown in **Figure 1C**. Shapiro Wilk tests showed that the distribution of hit rates did not significantly differ from normal in either session, all Ws > 0.975, all ps > .500. The participants' hit rates were submitted to a mixed-model ANOVA with *Session* as within-subjects factor and *Condition* as between-subjects factor. The main effect of Session was significant, $F(1, 109) = 303.479$, $MSE = 46.729$, $p < .001$, $\eta_p^2 = .736$, indicating higher hit rates in session 2 than in session 1. The main effect of Condition, $F(1, 109) = 0.620$, $MSE = 531.789$, $p = .433$, $\eta_p^2 = .006$, and the Session × Condition interaction, $F(1, 109) = 2.453$, $MSE = 46.729$, $p = .120$, $\eta_p^2 = .022$, were not significant.

Error scores. The percentages of false alarms and the percentages of misses were separately analyzed. Shapiro Wilk tests showed that the distribution of all error percentages deviated significantly from normal, all Ws < .96, all ps < .030. Therefore, we analyzed error percentages with nonparametric tests. Wilcoxon tests showed that both false alarms, $W = 1472.000$, $p < .001$, $r_b = .721$, and misses, $W = 4692.000$, $p < .001$, $r_b = .510$, decreased from session 1 to session 2. The decrease of false alarms in the position-repetition group (mean decrease = 0.162) did not differ from that in the position-change group (mean of decrease = 0.170), Mann-Whitney's $U = 1420.000$, $p = .746$, $r_b = .069$. Similarly, the decrease in misses in the position-repetition group (mean of decrease = 0.628) did not differ from that in the position-change group (mean decrease = 1.157), Mann-Whitney's $U = 1357.000$, $p = .841$, $r_b = .110$.

Discussion

Experiment 3 investigated whether repeating stimulus configurations (i.e., rows of stimuli) within and between tests contributes to practice effects in a d2-like test. Therefore, we compared changes in performance between two sessions in a position-repetition condition and a position-change condition. Most importantly, we observed similar improvements from session 1 to session 2 in all measures of performance in both conditions. Numerically, improvements were even larger in the position-change condition than in the position-repetition condition. Hence, Experiment 3 failed to show evidence that repeating stimulus rows in a d2-like test enables learning of stimulus configuration to a degree that could improve performance when the test is repeated once.

General Discussion

The main purpose of the present work was to investigate which components of a pen-and-paper test of attention (e.g., d2) are learned in a first session, and may therefore improve performance in a second session when the learned component is repeated. We conducted three experiments

Table 3. Summary statistics for performance scores observed in Experiment 3 as a function of Session (1, 2), and condition (position repetition, position change).

	Hits		False Alarms		Misses	
	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2
Position-repetition	84.880 (20.233)	99.500 (23.987)	0.244	0.082	3.799	3.172
Position-change	85.885 (20.596)	103.393 (24.861)	0.233	0.063	3.801	2.644
Total	85.383 (20.432)	101.447 (24.466)	0.238	0.072	3.800	2.882

Hits are given in absolute numbers, and also as percentages (in brackets) in relation to the total number of inspected items in session 1. False alarms and misses are given as percentages in relation to the total number of inspected items in the corresponding session.

with different, custom-made variants of the d2 test, in order to address this issue. In Experiment 1, we addressed learning of targets by comparing performance changes in a condition where targets were repeated to a condition where targets were changed between sessions. In Experiment 2, we addressed learning of distractors by comparing performance changes in a condition where distractors were repeated to a condition where distractors were changed between sessions. Finally, in Experiment 3, we addressed learning of stimulus configurations (positions) by comparing performance in a condition where configurations were repeated to a condition where configurations were changed between sessions.

The most important results of the experiments can be summarized as follows. In all three experiments, we observed significant improvements from session 1 to session 2 in all measures and conditions. In addition, the improvements were larger in Experiment 1 when targets repeated than when they changed, and in Experiment 2 when distractors repeated than when they changed. In particular, in Experiment 1, we observed a larger improvement in hits (KL) when targets repeated than when they changed. In Experiment 2, we observed a larger improvement (i.e., decrement) in false-alarms when distractors repeated than when they changed. In contrast, in Experiment 3 we observed similar improvements in performance when stimulus configurations repeated and when they changed.

The results of the present experiments provide strong evidence that test-takers practice and learn the detection and processing of targets during the first session with a d2-like test, and this learning improves performance in a second session when targets repeat (Experiment 1). Moreover, the results also provide moderate evidence that test-takers practice and learn the rejection of non-targets (i.e., distractors) during the first session with a d2-like test, and this learning also improves performance in a second session when distractors repeat (Experiment 2). Finally, the negative results of Experiment 3 suggest that repeating each

stimulus row several (i.e., five) times in the first session is not enough practice for learning these configurations, and therefore repeating the configurations in the second session has no measurable impact on performance. In the following sections, we will discuss practical and theoretical implications of our findings.

Practical implications

A major goal of the present work was to gain advice from experimental results for the construction of new attention tests that are less vulnerable to practice effects than are existing tests. The results of our experiments show that repeating targets within and between tests makes a strong contribution to practice benefits, whereas repeating distractors had a smaller impact. This pattern of findings implies that practice benefits from the repetition of d2-like tests of attention could be substantially reduced when the frequent repetition of targets is avoided. We will describe two possibilities for constructing new tests that avoid the frequent repetition of targets.

The first possibility is to construct tests in which the targets change regularly, while a consistent mapping of stimuli onto roles (target or distractor) is maintained. A large number of stimuli (e.g., 20) would be required for constructing a test like this.[3] The larger part of stimuli (i.e., 14) would become targets, whereas the rest of the stimuli (i.e., 4-6) would become distractors. The main difference to existing tests, such as the d2 or the FAIR, would be that participants had to search for a new target in each line of the test. Therefore, the to-be-searched target would be presented at the beginning of each line. Hence, 14 different targets would be required for building a test with 14 lines. Although participants are searching for only one target in each line, the test might have a sufficient level of difficulty when unfamiliar stimuli are used, and the target changes frequently.

---

3 This conclusion matches a suggestion made by Schumann, Steinborn, et al. (2022, p. 5) that increasing item set provides a means for mitigating practice gains in Düker-type tests of cognitive performance.

The second possibility of avoiding practice gains from the frequent repetition of the same targets is to use a variable stimulus-role mapping instead of using a consistent mapping. When using a variable mapping, a small number of stimuli (i.e., four) could be used for constructing the test because targets and distractors changed roles after each line. Hence, in odd-numbered lines, stimuli 1 and 2 were targets, and stimuli 3 and 4 were distractors, whereas the opposite mapping would apply in even-numbered lines. Previous studies on the effects of mapping on practice in visual-search tasks have shown that variable mapping significantly reduces practice effects in performance, when compared to consistent mapping (e.g., Fisk et al., 1991; Rogers, 1992; Rogers & Fisk, 1991; Schneider & Shiffrin, 1977), and therefore we expect that variable mapping will also reduce stimulus-related practice effects in pen-and-paper tests of attention.

Theoretical aspects

Although testing between theoretical accounts of practice effects in visual search was not an aim of the present experiments, we will nevertheless discuss accounts of our results. In fact, activation-based theories and retrieval-based theories can account for the main results of our experiments. Activation-based theories assume that selective attention leads to different activation levels of the cognitive representations of targets and distractors, and the different activation levels can have short-term and long-term effects on stimulus processing. In their activation-strength theory of automatization in visual search, Schneider and Shiffrin (1977) assume that sufficient practice in a visual-search task with consistent mapping produces durable changes in the attentional weights attached to representations of target and distractor stimuli. In particular, practice increases the attentional weights of targets that therefore attain the ability to automatically attract attention. In contrast, practice decreases the attentional weights of distractors that therefore attain the ability to automatically repel attention (e.g., Rogers, 1992; Schneider & Shiffrin, 1977; Shiffrin, 1988). Activation-based theories could also explain stronger learning of targets as compared to distractors, as suggested in our data, by referring to different levels of processing (e.g., Craik & Lockhart, 1972; Craik & Tulving, 1975). In particular, one might assume that targets are processed more deeply than distractors, and therefore the change of the attentional weights of targets is larger than the change of attentional weights of distractors.

Retrieval-based theories assume that participants store stimulus-response episodes in memory, and the memorized episodes subsequently affect the processing of similar stimuli (e.g., Logan, 1988, 1990; Neill, 1997, 2007). Instance theory of automatization (Logan, 1988, 1990) is a retrieval-based theory that has been proposed to explain practice-related improvements in performance and the automatization of skill. According to this theory, every stimulus-response episode is stored as a separate memory trace (called "instance") in memory. Moreover, the theory assumes that the response to a stimulus can either be computed by algorithmic processes, or can be retrieved from a memory trace of a previous encounter with that stimulus. Perceiving a stimulus triggers the retrieval of all similar instances in memory. The more similar instances containing a stimulus are retrieved upon presentation of that stimulus, the more likely an associated response will be retrieved from memory, and the faster (and more accurate) this response will be. Instance theory can explain stimulus-related practice effects in visual-search tasks, and in pen-and-paper tests of attention as the d2. Moreover, retrieval-based theories could also explain stronger learning of targets as compared to distractors, by referring to different levels of processing for targets and distractors (e.g., Craik & Tulving, 1975).

Activation-based and retrieval-based theories can equally well explain short-term priming effects (e.g., Fox, 1995; Frings et al., 2015). Moreover, there is additional support for each of these theories, and therefore many authors believe that both mechanisms contribute to priming effects (e.g., Frings et al., 2015; Neill, 2007; Tipper, 2001). In contrast, there is a lack of empirical studies on the contributions of activation-based or retrieval-based mechanisms to (long-term) practice effects in visual search, but instance theory (e.g., Logan, 1988, 1990) has received more empirical support as an explanation of automatization and skill acquisition than activation-based theories (e.g., Schneider & Shiffrin, 1977; Shiffrin, 1988).

Stimulus-independent sources of practice benefits

An interesting finding of our experiments is the observation of significant practice benefits in all change conditions. In particular, in Experiment 1, performance also improved from session 1 to session 2 in the target-change condition (e.g., P. Wühr, 2019). In the target-change condition, only distractors and stimulus configurations repeated, but Experiments 2 and 3 showed that these components alone produce only small practice benefits. Hence, the significant practice benefits in all measures in the target-change condition of Experiment 1 suggest that, in addition to stimulus-dependent (i.e., target and distractor) learning, stimulus-independent learning in the first session may also contribute to practice benefits. Stimulus-independent learning could be shown in a condition where all stimulus-related components (i.e., targets, distractors, and stimulus configurations) change, and only the task repeats. In fact, Wühr and Wühr (2021) report the results of such an experiment with a variant of the FAIR-2 test and observed significant improvements in performance from session 1 to session 2. Presumably, motor learning, increasing familiarity with task and test situation, and probably other variables, contribute to stimulus-independent learning. The existence of stimulus-independent learning implies that practice benefits will still occur in tests in which stimulus-dependent practice effects are prevented, although these residual practice benefits are certainly than in a complete-repetition condition.

Methodological issues

A first methodological issue concerns the basic design used for isolating the effects of target practice, distractor practice and configuration practice. We chose a design in which a "repeat everything" condition is compared to a "repeat everything but x" condition in order to isolate practice effects for targets, distractors and stimulus configurations. We had two major reasons for using that design. Firstly, the usual situation in assessment practice is the "repeat everything" condition, for which the amount of practice benefits is known. Therefore, we found it straightforward to analyze the effects of changing task components in comparison to this "repeat everything" condition. Secondly, one aim of our work is to identify ways for reducing practice effects in d2-like tests, and our design allows for directly testing which manipulations reduce practice effects. An alternative design for our experiments would have compared a "change everything but x" condition to a "change everything" condition. Interestingly, Le Dantec et al. (2012) observed comparable practice effects when "repeat targets or distractors" conditions were compared to a "repeat targets and distractors" condition, and when "repeat targets or distractors" conditions were compared to a "change targets and distractors" condition (the practice effects in RTs were computed from Table 1 in Le Dantec et al., 2012). These findings suggest that the two research designs can reveal similar estimates of practice effects.

A second methodological issue concerns possible effects of demand characteristics in our experiments. In Experiment 1, participants of both experimental conditions were tested together, and all participants were informed about the experimental manipulation in session 2. In particular, when being instructed for session 2, participants of Experiment 1 were told that some participants would have to search for the same targets as in the preceding session, whereas other participants would have to search for a new set of targets. Although being unavoidable, informing participants about the experimental manipulation might have affected the results of Experiment 1. For example, some participants might have guessed the hypotheses and performed accordingly. In that case, participants in the target-repetition condition might have spent more effort than participants in the target-change condition. Since participants in both conditions were informed about the experimental manipulation, a reverse effect is also possible. In that case, participants in the target-change condition might have spent more effort than participants in the target-repetition condition, because the former participants were aware of being in a presumably more difficult condition. Since we have no means of determining whether and how demand characteristics affected the results of Experiment 1, the possibility of such effects should be kept in mind when interpreting these results. Demand characteristics should not have affected the results of Experiments 2 and 3 because participants in these two experiments were not informed about the experimental manipulations (i.e., differences in stimulus materials between sessions).

A third methodological issue concerns the possible role of stimulus similarity. In the target-change condition of Experiment 1, the target set changed from [,d' − ,,d − d''] to [,p' − ,,p − p'']. In the distractor-change condition of Experiment 2, the distractor set changed from [,d − ,,d' − ,p' − ,,p − p''] to [d' − ,d'' − ,q' − ,,q − q'']. Hence, in both cases, the letters of three stimulus items changed from session 1 to session 2. It is, however, possible that replacing d with p, in the target set of Experiment 1, produced a bigger difference than replacing p with q, in the distractor set of Experiment 2 (e.g., Boles & Clifford, 1989). As a result, the targets in the two sessions of the target-change condition of Experiment 1 would have been less similar than the distractors in the two sessions of the distractor-change condition of Experiment 2. If this was correct, less similarity of targets in subsequent sessions of Experiment 1 could have allowed for less transfer of target learning, and thus produced a relatively large performance difference between conditions in the second session of Experiment 1. In contrast, more similarity of distractors in subsequent sessions of Experiment 2 could have allowed for more transfer of distractor learning, and thus produced a relatively small performance difference between conditions in the second session of Experiment 2. Hence, comparing the size of practice benefits between these experiments might be compromised by different changes of stimulus similarities in the critical change conditions. A larger difference between distractor sets in the distractor-change condition might have produced a larger distractor-related practice effect. In any case, the fact that repeating distractors produced better performance than changing distractors shows that distractor learning can contribute to practice benefits in d2-like tests.

Conclusion

In a nutshell, the results suggest that target learning makes a strong contribution, distractor learning makes a moderate contribution, and contextual learning makes a negligible contribution to the practice gains that are observed when a d2-like test is repeated. These findings might be considered when constructing new tests of focused attention that are less vulnerable to practice benefits.

Funding Information

Competing Interests

We do not have competing interests to report.

Data Accessibility Statement

The raw data reported in this article (paper sheets) cannot be made available on a public repository. However, readers may obtain excel sheets containing test scores for each participant in each experiment from the authors upon request.

References

Blotenberg, I., & Schmidt-Atzert, L. (2019). On the locus of the practice effect in sustained attention tests. *Journal of Intelligence*, *7*(2), 12. https://doi.org/10.3390/jintelligence7020012

Blotenberg, I., & Schmidt-Atzert, L. (2020). On the characteristics of sustained attention test performance—The role of the preview benefit. *European Journal of Psychological Assessment*, *36*(4), 593–600. https://doi.org/10.1027/1015-5759/a000543

Boles, D. B., & Clifford, J. E. (1989). An upper- and lowercase alphabetic similarity matrix, with derived generation similarity values. *Behavior Research Methods, Instruments, & Computers*, *21*(6), 579–586. https://doi.org/10.3758/bf03210580

Brickenkamp, R. (1962). *Aufmerksamkeits-Belastungs-Test d-2 [Sustained-attention test d-2]*. Hogrefe.

Brickenkamp, R. (2002). *Test d2 – Revision*. Hogrefe.

Brickenkamp, R., Schmidt-Atzert, L., & Liepmann, D. (2010). *Test d2 – Revision (d2-R)*. Hogrefe.

Brickenkamp, R., & Zillmer, E. (1998). *D2 Test of Attention*. Hogrefe.

Chan, L. K. H., & Hayward, W. G. (2013). Visual search. *WIREs Cognitive Science*, *4*(4), 415–429. https://doi.org/10.1002/wcs.1235

Chao, H.-F., & Yeh, Y.-Y. (2005). Location negative priming in identity discrimination relies on location repetition. *Perception & Psychophysics*, *67*(5), 789–801. https://doi.org/10.3758/bf03193533

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, *4*(5), 170–178. https://doi.org/10.1016/s1364-6613(00)01476-5

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/s0022-5371(72)80001-x

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Czerwinski, M., Lightfoot, N., & Shiffrin, R. M. (1992). Automatization and training in visual search. *The American Journal of Psychology*, *105*(2), 271–315. https://doi.org/10.2307/1423030

Fisk, A. D., Lee, M. D., & Rogers, W. A. (1991). Recombination of automatic processing components: The effects of transfer, reversal, and conflict situations. *Human Factors*, *33*(3), 267–280. https://doi.org/10.1177/001872089103300303

Fox, E. (1995). Negative priming from ignored distractors in visual selection: A review. *Psychonomic Bulletin & Review*, *2*(2), 145–173. https://doi.org/10.3758/bf03210958

Frings, C., Schneider, K. K., & Fox, E. (2015). The negative priming paradigm: An update and implications for selective attention. *Psychonomic Bulletin & Review*, *22*(6), 1577–1597. https://doi.org/10.3758/s13423-015-0841-4

Frings, C., & Wühr, P. (2007a). On distractor-repetition benefits in the negative-priming paradigm. *Visual Cognition*, *15*(2), 166–178. https://doi.org/10.1080/13506280500475264

Frings, C., & Wühr, P. (2007b). Prime display offset modulates negative priming only for easy-selection tasks. *Memory & Cognition*, *35*(3), 504–513. https://doi.org/10.3758/bf03193290

Geng, J. J., Won, B.-Y., & Carlisle, N. B. (2019). Distractor ignoring: Strategies, learning, and passive filtering. *Current Directions in Psychological Science*, *28*(6), 600–606. https://doi.org/10.1177/0963721419867099

Guy, S., & Buckolz, E. (2007). The locus and modulation of the location negative priming effect. *Psychological Research*, *71*(2), 178–191. https://doi.org/10.1007/s00426-005-0003-9

Hagemeister, C. (2007). How useful is the power law of practice for recognizing practice in concentration tests? *European Journal of Psychological Assessment*, *23*(3), 157–165. https://doi.org/10.1027/1015-5759.23.3.157

Hagemeister, C., Scholz, A., & Westhoff, K. (2002). Wie kann man Geübtheit in Konzentrationstests erkennbar machen? [How can practice be identified in concentration tests?]. *Zeitschrift für Personalpsychologie*, *1*(2), 94–102. https://doi.org/10.1026//1617-6391.1.2.94

Hagemeister, C., & Westhoff, K. (2011). Konzentrationsdiagnostik. In L. F. Hornke, M. Amelang, & M. Kersting (Eds.), *Leistungs-, Intelligenz- und Verhaltensdiagnostik* (pp. 51–96). Hogrefe.

Harris, A. M., & Remington, R. W. (2017). Contextual cueing improves attentional guidance, even when guidance is supposedly optimal. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(5), 926–940. https://doi.org/10.1037/xhp0000394

Harris, J. G., Minassian, A., & Perry, W. (2007). Stability of attention deficits in schizophrenia. *Schizophrenia Research*, *91*(1–3), 107–111. https://doi.org/10.1016/j.schres.2006.12.021

Hemmerich, W. (2015–2022). *StatistikGuru (version 1.96)*. https://statistikguru.de

Houghton, G., & Tipper, S. P. (1994). A model of inhibitory mechanisms in selective attention. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 53–112). Academic Press.

Huang, L., Holcombe, A. O., & Pashler, H. (2004). Repetition priming in visual search: Episodic retrieval, not feature priming. *Memory & Cognition*, *32*(1), 12–20. https://doi.org/10.3758/bf03195816

Le Dantec, C. C., Melton, E. E., & Seitz, A. R. (2012). A triple dissociation between learning of target, distractors, and spatial contexts. *Journal of Vision*, *12*(2), 5–5. https://doi.org/10.1167/12.2.5

Lievens, F., Reeve, C. L., & Heggestad, E. D. (2007). An examination of psychometric bias due to retesting on cognitive ability tests in selection settings. *Journal of Applied Psychology*, *92*(6), 1672–1682. https://doi.org/10.1037/0021-9010.92.6.1672

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527. https://doi.org/10.1037/0033-295x.95.4.492

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, *22*(1), 1–35. https://doi.org/10.1016/0010-0285(90)90002-l

Makovski, T. (2016). What is the context of contextual cueing? *Psychonomic Bulletin & Review*, *23*(6), 1982–1988. https://doi.org/10.3758/s13423-016-1058-x

Moosbrugger, H., & Oehlschlägel, J. (2011). *Frankfurter Aufmerksamkeits-Inventar 2 (FAIR-2)*. Verlag Hans Huber.

Neill, W. T. (1977). Inhibitory and facilitatory processes in selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(3), 444–450. https://doi.org/10.1037/0096-1523.3.3.444

Neill, W. T. (1997). Episodic retrieval in negative priming and repetition priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(6), 1291–3105. https://doi.org/10.1037/0278-7393.23.6.1291

Neill, W. T. (2007). Mechanisms of transfer-inappropriate processing. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 63–78). American Psychological Association. https://doi.org/10.1037/11587-004

Neill, W. T., & Valdes, L. A. (1992). Persistence of negative priming: Steady state or decay? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 565–576. https://doi.org/10.1037/0278-7393.18.3.565

Neisser, U. (1963). Decision-time without reaction-time: Experiments in visual scanning. *The American Journal of Psychology*, *76*(3), 376–385. https://doi.org/10.2307/1419778

Neumann, E., & DeSchepper, B. G. (1991). Costs and benefits of target activation and distractor inhibition in selective attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(6), 1136–1145. https://doi.org/10.1037/0278-7393.17.6.1136

Parasuraman, R., & Davies, D. R. (1984). *Varieties of attention*. Academic Press.

Park, J., & Kanwisher, N. (1994). Negative priming for spatial locations: Identity mismatching, not distractor inhibition. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(3), 613–623. https://doi.org/10.1037/0096-1523.20.3.613

Pashler, H. E. (1998). *The psychology of attention*. The MIT Press.

Prinz, W. (1979). Locus of the effect of specific practice in continuous visual search. *Perception & Psychophysics*, *25*(2), 137–142. https://doi.org/10.3758/bf03198800

Rivera, D., Salinas, C., Ramos-Usuga, D., Delgado-Mejía, I. D., Vasallo Key, Y., Hernández Agurcia, G. P., Valencia Vásquez, J., García-Guerrero, C. E., García de la Cadena, C., Rabago Barajas, B. V., Romero-García, I., Campos Varillas, A. I., Sánchez-SanSegundo, M., Galvao-Carmona, A., Lara, L., Granja Gilbert, E. J., Martín-Lobo, P., Velázquez-Cardoso, J., Caracuel, A., & Arango-Lasprilla, J. C. (2017). Concentration Endurance Test (d2): Normative data for Spanish-speaking pediatric population. *NeuroRehabilitation*, *41*(3), 661–671. https://doi.org/10.3233/nre-172248

Rogers, W. A. (1992). Age differences in visual search: Target and distractor learning. *Psychology and Aging*, *7*(4), 526–535. https://doi.org/10.1037/0882-7974.7.4.526

Rogers, W. A., & Fisk, A. D. (1991). Are age differences in consistent-mapping visual search due to feature learning or attention training? *Psychology and Aging*, *6*(4), 542–550. https://doi.org/10.1037/0882-7974.6.4.542

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 1–17. https://doi.org/10.1037/0096-1523.3.1.1

Schmidt-Atzert, L., Büttner, G., & Bühner, M. (2004). Theoretische Aspekte von Aufmerksamkeits-/Konzentrationsdiagnostik [Theoretical aspects of sustained-attention tests]. In G. Büttner & L. Schmidt-Atzert (Eds.), *Diagnostik von Konzentration und Aufmerksamkeit* (pp. 3–22). Hogrefe.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. https://doi.org/10.1037/0033-295x.84.1.1

Schumann, F., Steinborn, M. B., Flehmig, H. C., Kürten, J., Langner, R., & Huestegge, L. (2022). On doing multi-act arithmetic: A multitrait-multimethod approach of performance dimensions in integrated multitasking. *Frontiers in Psychology*, *13*, 946626. https://doi.org/10.3389/fpsyg.2022.946626

Shiffrin, R. M. (1988). Attention. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology: Perception and motivation; Learning and cognition* (2nd ed., Vols. 1–2, pp. 739–811). John Wiley & Sons.

Sisk, C. A., Remington, R. W., & Jiang, Y. V. (2019). Mechanisms of contextual cueing: A tutorial review. *Attention, Perception, & Psychophysics*, *81*(8), 2571–2589. https://doi.org/10.3758/s13414-019-01832-2

Steinborn, M. B., Langner, R., Flehmig, H. C., & Huestegge, L. (2018). Methodology of performance scoring in the d2 sustained-attention test: Cumulative-reliability functions and practical guidelines. *Psychological Assessment*, *30*(3), 339–357. https://doi.org/10.1037/pas0000482

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *37A*(4), 571–590. https://doi.org/10.1080/14640748508400920

Tipper, S. P. (2001). Does negative priming reflect inhibitory mechanisms? A review and integration of conflicting views. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *54A*(2), 321–343. https://doi.org/10.1080/02724980042000183

Tipper, S. P., Bourque, T. A., Anderson, S. H., & Brehaut, J. C. (1989). Mechanisms of attention: A developmental study. *Journal of Experimental Child Psychology*, *48*(3), 353–378. https://doi.org/10.1016/0022-0965(89)90047-7

Tipper, S. P., & Cranston, M. (1985). Selective attention and priming: Inhibitory and facilitatory effects of ignored primes. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *37A*(4), 591–611. https://doi.org/10.1080/14640748508400921

Treisman, A. M. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *40A*(2), 201–237. https://doi.org/10.1080/02724988843000104

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

Westhoff, K., & Dewald, D. (1990). Effekte der Übung in der Bearbeitung von Konzentrationstests [The effects of practice on concentration-test performance]. *Diagnostica*, *36*, 1–15.

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. https://doi.org/10.3758/bf03200774

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–73). Psychology Press/Erlbaum (UK) Taylor & Francis.

Wühr, B., & Wühr, P. (2021). Effects of repeated testing in a pen-and-paper test of selective attention (FAIR-2). *Psychological Research*, *86*(1), 294–311. https://doi.org/10.1007/s00426-021-01481-x

Wühr, P. (2019). Target-specific learning contributes to practice effects in paper-and-pencil tests of attention. *Swiss Journal of Psychology*, *78*(1–2), 29–35. https://doi.org/10.1024/1421-0185/a000221

Yato, Y., Hirose, S., Wallon, P., Mesmin, C., & Jobert, M. (2019). d2-R test for Japanese adolescents: Concurrent validity with the attention deficit-hyperactivity disorder rating scale. *Pediatrics International*, *61*(1), 43–48. https://doi.org/10.1111/ped.13735

Supplementary Materials

Peer Review History

Download: https://collabra.scholasticahq.com/article/71297-disentangling-the-contributions-of-repeating-targets-distractors-and-stimulus-positions-to-practice-benefits-in-d2-like-tests-of-attention/attachment/148388.docx?auth_token=OvJwLNEQ40t5gAhTWWoJ

Appendix

Download: https://collabra.scholasticahq.com/article/71297-disentangling-the-contributions-of-repeating-targets-distractors-and-stimulus-positions-to-practice-benefits-in-d2-like-tests-of-attention/attachment/148390.docx?auth_token=OvJwLNEQ40t5gAhTWWoJ