Research Article

Spatial neglect in the digital age: Influence of presentation format on patients' test behavior

Hannah Rosenzopf[1] ⓘ, Christoph Sperber[1] ⓘ, Franz Wortha[2], Daniel Wiesen[1], Annika Muth[1], Elise Klein[3,4], Korbinian Möller[4,5,6] and Hans-Otto Karnath[1,7]

[1]Centre of Neurology, Division of Neuropsychology, Hertie-Institute for Clinical Brain Research, University of Tuebingen, Tuebingen, Germany, [2]Department of Psychology, University of Greifswald, Greifswald, Germany, [3]University of Paris, LaPsyDÉ, CNRS, Sorbonne Paris Cité, Paris, France, [4]Leibniz Institut für Wissensmedien, Tuebingen, Germany, [5]Centre for Mathematical Cognition, School of Science, Loughborough University, Loughborough, United Kingdom, [6]Centre for Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt, Germany and [7]Department of Psychology, University of South Carolina, Columbia, SC, USA

Abstract

Objective: Computerized neglect tests could significantly deepen our disorder-specific knowledge by effortlessly providing additional behavioral markers that are hardly or not extractable from existing paper-and-pencil versions. This study investigated how testing format (paper versus digital), and screen size (small, medium, large) affect the Center of cancelation (CoC) in right-hemispheric stroke patients in the Letters and the Bells cancelation task. Our second objective was to determine whether a machine learning approach could reliably classify patients with and without neglect based on their search speed, search distance, and search strategy. **Method:** We compared the CoC measure of right hemisphere stroke patients with neglect in two cancelation tasks across different formats and display sizes. In addition, we evaluated whether three additional parameters of search behavior that became available through digitization are neglect-specific behavioral markers. **Results:** Patients' CoC was not affected by test format or screen size. Additional search parameters demonstrated lower search speed, increased search distance, and a more strategic search for neglect patients than for control patients without neglect. **Conclusion:** The CoC seems robust to both test digitization and display size adaptations. Machine learning classification based on the additional variables derived from computerized tests succeeded in distinguishing stroke patients with spatial neglect from those without. The investigated additional variables have the potential to aid in neglect diagnosis, in particular when the CoC cannot be validly assessed (e.g., when the test is not performed to completion).

Keywords: diagnostics; test digitalization; center of cancelation; right hemisphere; stroke; Human

Introduction

Spatial neglect is a common result of unilateral, predominantly right-hemispheric brain damage (Becker & Karnath, 2007). Its core symptoms include an egocentric bias in gaze direction and exploration towards the ipsilesional side (Corbetta & Shulman, 2011; Karnath & Rorden, 2012). One type of test to detect and quantify these symptoms are cancelation tasks (Weintraub & Mesulam, 1985; Gauthier, Dehaut & Joanette, 1989; Ferber & Karnath, 2001). They are commonly presented on sheets of paper placed in front of the patient, who is required to find and manually mark all targets among distractors. Patients with spatial neglect often miss targets on the contralesional side. The presence and severity of spatial neglect can be measured by the Center of Cancelation (CoC, Rorden & Karnath, 2010) which assesses the average position of correctly marked targets with respect to the patient's ego center.

While paper-and-pencil-based cancelation tasks can be a time-efficient yet reliable diagnostic alternative to more extensive test batteries (Fullerton, Stout, & McSherry, 1986; Ferber & Karnath, 2001), they provide only part of the information they could if they were computer-based (Schendel & Robertson, 2002; Bonato & Deouell, 2013, Dalmaijer, Van der Stigchel, Nijboer, Cornelissen, & Husain, 2015). Among other aspects, digitization can provide additional variables such as response time, revisits (of already marked items), and information concerning the search path applied (Donnelly et al., 1999; Dalmaijer et al., 2015).

However, due to the lack of comparison with traditional, validated paper-and-pencil versions, it cannot yet be excluded that variations in test format may lead to results that differ from those of traditional paper-and-pencil versions. Furthermore, in clinical practice, traditional A4 paper-and-pencil tests will likely be implemented as scaled-down versions matching commonly used tablet sizes. However, the effect of using devices of different sizes on the validity of the tests has not yet been sufficiently studied in the context of cancelation tasks. Concerning line bisection, another means used to diagnose neglect, previous observations have suggested that

the length of the bisected line may have some influence on spatial attentional processing (Bowers & Heilman, 1980; McCourt & Jewell, 1999; Anderson, 1997). On the other hand, studies in neurological patients have suggested that a change in frame size, that is the size of the space searched by the patient, does not necessarily affect neglect-specific impairments. Body-centered (egocentric) and object-centered (allocentric) neglect appeared to dynamically adapt to different frame sizes, (Karnath & Niemeier, 2002; Baylis, Baylis, & Gore, 2004; Karnath, Mandler & Clavagnier, 2011; Li, Karnath & Rorden, 2014).

In the present study, we compared right hemisphere stroke patients' performance in cancelation tasks across different formats (paper-and-pencil vs. digital) and display sizes (small, medium, large) to investigate whether digitization of traditional cancelation tasks to various screen sizes affects their validity. As new variables become available through digitization, a further objective was to evaluate their contribution to diagnostic decisions. This is important because in clinical practice patients not always can complete a cancelation task (e.g., because they are too exhausted or because testing must be interrupted due to other clinical necessities). While measuring the CoC requires running the test to completion, other behavioral variables might become extractable already early on and thus aid diagnosis (if a test cannot be completed), given that these parameters proved to detect neglect-specific behavior. Based on previous observations on neglect patients' visual coordination (Karnath & Huber, 1992; Donnelly et al., 1999; Ptak, Golay, Müri, & Schnider, 2009; Machner et al., 2012; Kaufmann et al., 2020) we investigated parameters search speed (number of targets found relative to time), search distance (the mean distance between two consecutive targets), and search strategy (a calculation of search path) for their ability to predict spatial neglect.

Methodological investigations have shown that effects revealed by statistical analyses often have limited informative value in (applied) diagnostic contexts, even when effect sizes are very large (Dwyer, Falkai & Koutsouleris, 2018). Due to their strong focus on generalization and prediction of unknown data, machine learning approaches appear to be more suitable in most diagnostic applications than most statistical modeling approaches (Dwyer, Falkai & Koutsouleris, 2018). The specific use of machine learning models in diagnostic processes can vary, ranging from automatic evaluations of diagnostic tasks (Chen et al., 2020) to interpretable classifications that outperform traditional paper-based tests in the prediction of neuropsychiatric disorders (Souillard-Mandar et al., 2021). Accordingly, in the present investigation, we tested the potential diagnostic value of process parameters obtained from digital cancelation tests using such approaches.

Method

Subjects

Nineteen continuously admitted acute right hemisphere stroke patients (N = 8 without spatial neglect; N = 11 suffering from spatial neglect) and one chronic neglect patient who returned for a follow-up neuropsychological investigation were recruited at the Centre of Neurology at Tuebingen University. Structural imaging was acquired by computed tomography as part of the clinical routine conducted for all stroke patients at admission except for one patient who received magnetic resonance imaging instead. Patients with diffuse or bilateral brain lesions, patients with tumors, and patients without obvious lesions were not included. According to the routine clinical neurological examination, patients did not suffer from any further neurological pathologies. Clinical and

Table 1. Demographic and clinical data of the 20 right-hemispheric patients with and without spatial neglect included in the study. Mean, standard deviation

	Neglect	No neglect
Age in years	71.0, 8.3	56.9, 15.9
Sex (M/F)	8/4	7/1
Etiology (Ischemia/Hemorrhage)	12/0	8/0
Visual Field Defects (% present)	25	12.5
Hemiparesis (% present)	100	78
Letters cancelation test		
CoC*	0.374, 0.295	0.0004, 0.018
Targets found L/R*	11/23	28/29
Revisits**	1.06, 1.58	0.59 0.66
False alarms**	0.09, 0.39	0.05, 0.21
Bells test		
CoC*	0.395, 0.235	0.004, 0.049
Targets found L/R*	4/13	13/13
Revisits**	0.75 1.76	0.21, 0.58
False alarms**	0, 0	0, 0
Copying task (N omitted)*	2.56	0.38

*derived from initial diagnostics; **pooled from digital versions (data was not evident from paper-and-pencil versions).

demographic variables of the two patient groups are summarized in Table 1; Figure 1 illustrates an overlap plot of their brain lesions. The study was performed in accordance with the revised Declaration of Helsinki, the local ethics committee approved the study and all patients provided their written consent to participate.

All patients were clinically examined with a bedside neglect screening upon admission to the Centre of Neurology. This screening determined patients' allocation to the neglect group or the control group. The 19 acute stroke patients were tested on average 6.4 days (SD = 4.5) post-stroke; the chronic neglect patient was tested 32 months post-stroke. The screening included two cancelation tasks (Bells test [Gauthier et al., 1989]; Letters test [Weintraub & Mesulam, 1985]), and a copying task (Johannsen & Karnath, 2004). These tasks were presented on a DIN A4 sized 297 by 210 mm paper each. We calculated the CoC using the procedure and cut-off scores for neglect diagnosis by Rorden and Karnath (2010) for both the Letters (cut-off: −/+ 0.083) and Bells test (cut-off: −/+ 0.081). The CoC is a sensitive measure capturing both number and location of omissions, with zero representing an equal distribution of correctly identified stimuli along the x-axis of the test sheet. Negative deviations (with a maximum of −1) indicate a bias to the left side of the test sheet. Positive deviations (with a maximum of 1) indicate a bias to the right side of the test sheet. The copying task requires patients to copy a complex multi-object scene consisting of four figures (a fence, a car, a house, and a tree), with two of them located in each half of the horizontally oriented sheet of paper. Omission of at least one of the contralesional features of each figure was scored as 1, and omission of each whole figure was scored as 2. One additional point was given when contralesionally located figures were drawn on the ipsilesional side of the paper sheet. The maximum score was 8. A score higher than 1 (i.e., > 12.5% omissions) was taken to indicate neglect. The duration of each test depended on the patient being satisfied with his/her performance and confirming this twice. Spatial neglect was diagnosed if patients scored within the pathological ranges of at least 2 out of 3 tests (see. Tab. 1).

Material and procedure

The experiment included the same cancelation tasks as the clinical assessment, that is, the Bells and Letters test, presented on
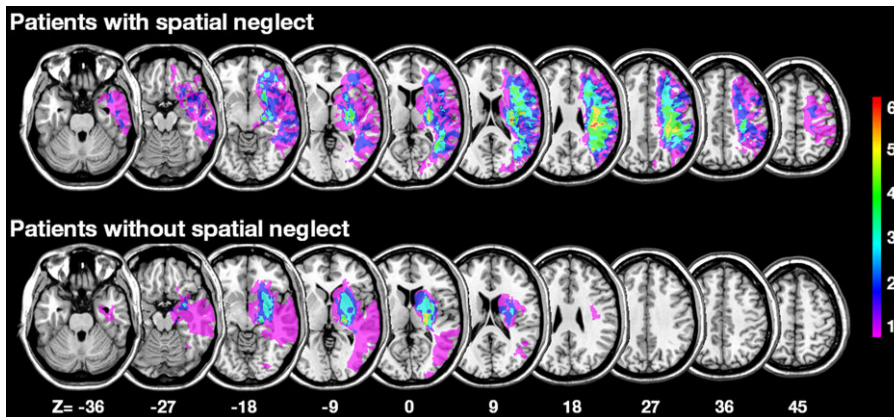
Figure 1. Lesion overlays. Overlay of the normalized lesions of right hemisphere patient groups with and without spatial neglect. Lesion boundaries were semi-automatically using the Clusterize algorithm on the SPM Clusterize toolbox (cf. De Haan et al., 2015) on SPM 12 (http://www.fil.ion.ucl.ac.uk/spm). Normalization of CT or MR scans to MNI space with 1x1x1 mm resolution was performed by using the Clinical Toolbox (Rorden, Hjaltason et al., 2012) under SPM12, and by registering lesions to its age-specific MR or CT templates oriented in MNI space (Rorden, Bonilha et al., 2012).

A4 sheets of paper. Beyond, the experiment comprised computerized touch screen versions of said cancelation tasks. Computerized testing was performed on a capacitive 27-inch multi-touch display (3M – M2767PW), connected to a laptop (HP ProBook 4740s with Windows 7 Professional). The touchscreen versions of the two tasks were custom created using MATLAB 2016b and Psychtoolbox (https://doi.org/10.17632/6dzxs69j7d.1). Computerized cancelation tests (touch screen – TS) were high-resolution versions of the original test images used for the paper-and-pencil version displayed in three different sizes: 260.28 mm × 173.52 mm ("TS small"; a tablet size as e.g., in Microsoft surface, HP Elite, Dell Latitude 5290), 297 mm × 210 mm ("TS medium"; equivalent to an A4 paper), and 597.6 mm × 336.2 mm ("TS large"; full-screen size of the 27-inch touch screen). The small and medium versions were displayed centrally on the 27-inch display, with a black margin between the end of the test and the end of the screen. Despite the different sizes in the respective conditions test coordinates were always measured with a relative distance from center to borders between −1 and 1, −1/−1 representing the upper left corner. To keep paper-and-pencil and touchscreen conditions as comparable as possible, the touchscreen lay flat on the table and a touchscreen compatible pen (Adonit Dash 2) was used to mark the targets. Patients' marks were visualized in real-time, providing patients with visual feedback comparable to that provided by conventional pens on a regular sheet of paper. Due to their health issues, four patients were unable to complete all trials, which led to 9 missing data sets in different test conditions. Said patients had to be excluded from parts of the analyses.

In the experiment, half of the participants started with the paper-and-pencil version of the two cancelation tasks, the other half with the touchpad versions. The order of the two paper-and-pencil versions was alternated, the order of the 6 different touchpad versions was randomized. Participants were instructed to find all the bells/"A"s that were spread among distractors and to tell the experimenter once they were done. Before starting the next trial, patients confirmed that they were indeed done with this trial, that is, could not find any other target stimuli.

Data analysis

For comparing right hemisphere stroke patients' CoC performance in cancelation tasks across different formats (paper-and-pencil vs. digital) and display sizes (small, medium, large), we used Wilcoxon and Friedman tests respectively. To measure (1) search speed, we extracted a participant's total number of correctly identified items and divided it by the time measured between starting the test and marking the last item to assess the number of targets found relative to time (measured in seconds). For (2) search distance we averaged the Euclidean distance between every two targets found in direct succession to each other. While search distance was defined as Euclidean distance, a high degree of (3) search strategy, was defined by a pattern that keeps either the steps along the (assumed) x- or y-axis low and subsequently results in a row (a low distance on the y-axis) or column-wise (a low distance on the x-axis) search (for an illustration of the distinction see Figure 2). Both distances were averaged across all found targets. A strategic search, as we define it here, should result in low values in either the mean x-axis distance or the mean y-axis distance. Low y values indicate a row-wise left-to-right (reading-like; Figure 3A) or alternating left-to-right and right-to-left (Figure 3B) search pattern; low x values indicate a column-wise top-to-bottom (Figure 3C) or alternating top-to-bottom and bottom-to-top (Figure 3D) search pattern. The measure is independent of direction and applies also if tests were started from the right or the bottom. To investigate potential differences between (i) the digital screen sizes and (ii) right-hemispheric patients with spatial neglect in comparison to patients without neglect, we applied a 2 × 3 analyses of variance for each of the three parameters above (i.e., search speed, search distance, and search strategy) with the between-subjects factor group (neglect vs. no neglect) and the within-subjects factor screen size (TS small vs. TS medium vs. TS large).

To finally analyze if the three parameters above can be used to reliably predict participants' neglect diagnosis (dichotomized: spatial neglect vs. no neglect), we used Support Vector Machines (SVM). Given that SVM require complete data sets, we first used Multiple Imputation by Chained Equations (MICE; White, Royston, & Wood, 2011) to impute missing data for this analysis step only. It entailed that missing values in a given column were estimated using a Bayesian Ridge Regressor, predicting values of the current column from all other columns. MICE was carried out column-wise from the column with the least number of missing values to the column with the most missing values. The potential impact of the imputation was tested by rerunning all analyses with a dataset where missing values were omitted. In the following sections, only results for the imputed dataset will be reported, because the pattern of results remained identical with and without imputation. Due to the sample size of the present study, we decided to use a dataset containing all screen sizes (TS small, TS medium, TS large) for each participant. To account for the dependence of

Figure 2. Distinction between distance and strategy. Distance was defined as the Euclidean distance between two consecutive targets. Search strategy in our case assumes that the more strategic the search is, the more it follows a row or column-wise search pattern, manifesting in small distances along the y- or x-axis, respectively. While it is possible that the target with the lowest Euclidean distance is also the most strategic one this is not necessarily the case. E.g. from position A target B minimizes both Euclidean distance and the distance along the y-axis. From position B, on the other hand, the most strategic step (i.e. minimizing y-distance as before) is target C, while the target that is overall the closest (and therefore minimizing Euclidean distance) is target D.
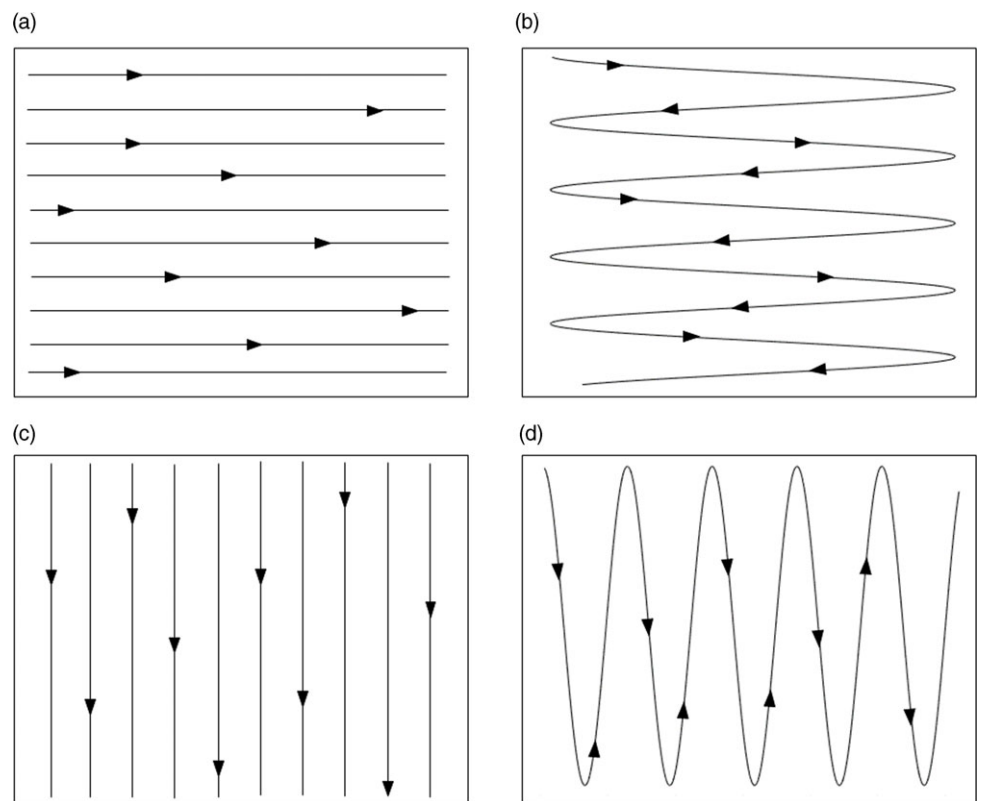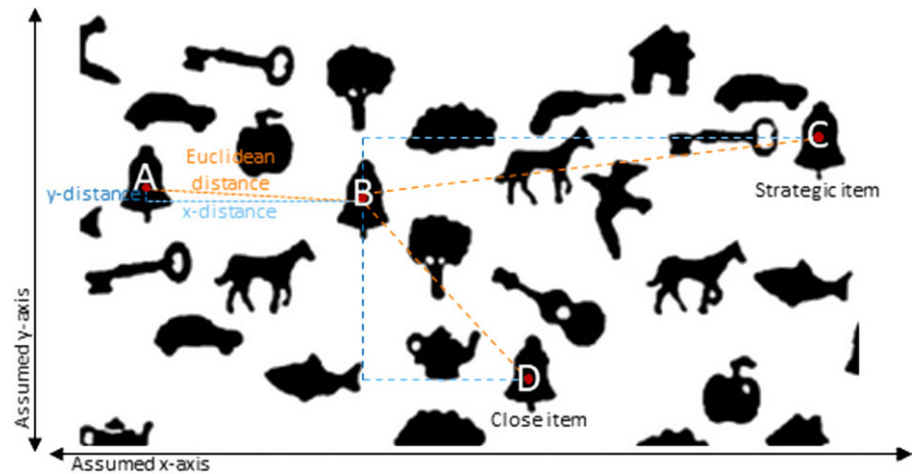

Figure 3. Measured search strategies. Search strategies covered by variable "search strategy" in the present study: (A) left-to-right (reading-like) strategy; (B) alternating left-to-right and right-to-left search pattern; (C) column-wise top-to-bottom strategy; (D) alternating top-to-bottom and bottom-to-top search pattern.

data points in this approach (i.e., three measures for each participant), we tested our models through Leave-One-Subject-Out Cross-Validation. In this procedure, the machine learning model is trained on data for all participants but one and tested on the participant that was left out for training. This process is then repeated until each participant was predicted once and prediction outcomes (i.e., balanced accuracy due to the unequal group sizes; Brodersen, Ong, Stephan & Buhmann, 2010) are averaged across predictions for all participants. Hyperparameters (i.e., the kernel: linear or radial basis function; cost parameter: ranging from 0.01 to 10) were optimized through a grid search in a nested Leave-One-Subject-Out Cross-Validation (within the training dataset). This procedure was carried out separately for each task (Bells and Letters test) and balanced accuracy scores were obtained across all screen sizes for

both tasks. Lastly, the percent of correctly classified neglect and right-hemispheric control patients were accumulated for each screen size and test. To test if the classification accuracy varied by screen size, chi-square tests of independence were used to compare the distribution of correctly classified neglect and right-hemispheric control patients across screen sizes for each task (Bells and Letters test). All machine learning analyses were conducted in Python using the scikit learn module (Pedregosa et al., 2011).

Results

Comparison between paper-and-pencil and digital formats

To investigate whether digital versus paper-and-pencil test format has an impact on patients' performance in cancelation tasks, we
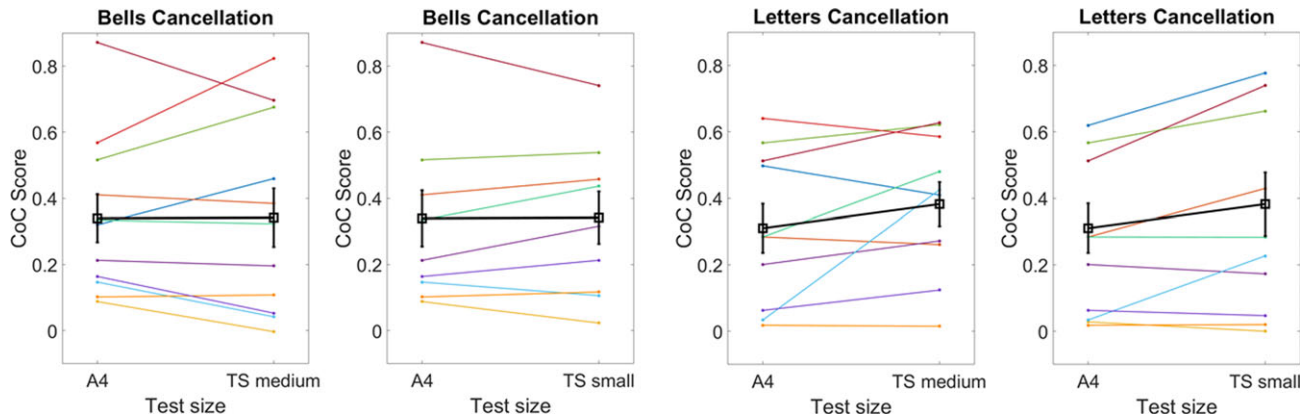
Figure 4. CoC in paper and pencil vs. digital versions. Neglect patients' performance for the Bells test and the Letter cancellation test in the traditional A4 paper-and-pencil version compared to the digital format with an equivalent touchscreen size (TS medium) as well as to the digital format with the smaller size that corresponds to a current tablet format (TS small). The bold lines represent mean values (including error bars) averaged over all patients.

compared patients' mean CoC scores in the A4 paper-and-pencil version to those in the same size in the digital A4 touch screen version (TS medium). Data are illustrated in Figure 4. Wilcoxon tests indicated no significant median CoC differences between the digital and the paper-and-pencil versions, neither in the Letters test ($Z = 1.784$, $p = 0.072$) nor in the Bells test ($Z = 0.533$, $p = 0.594$). In clinical practice, the traditional A4 paper-and-pencil tests will most likely be implemented as a down-scaled version to match the currently used tablet size. Thus, we also investigated (cf. Figure 4) whether differences in performance arise between the established A4 paper-and-pencil version and the digital downsized tablet size (TS small). Again, we did not find significant differences for neither the Letters test ($Z = 1.784$, $p = 0.074$) nor the Bells test ($Z = -0.356$, $p = 0.722$).

Comparison between different sizes of the digital format

Center of cancelation

To investigate whether size variation between the digital versions affects cancelation performances, we used the CoC as dependent variable and test size (TS small vs. TS medium vs. TS large) as independent variable. Data are illustrated in Figure 5. Friedman tests revealed no significant results for the Letters test ($\chi^2 F(2) = 1.750$, $p = 0.417$). The Bells test ($\chi^2 F(2) = 6.00$, $p = 0.050$) was right at the border to significance. We, therefore, applied post hoc Wilcoxon comparisons to rule out significant differences. Indeed all three were found to be non-significant.

Additional parameters of search behavior

Beyond the CoC, the additional variables, search speed, search distance, and search strategy were obtained from the digitized cancelation tasks.

Search speed. Data are illustrated in Figure 6. Analysis of the Bells test revealed a significant main effect of group ($F(1,15) = 6.719$, $p = 0.02$, $\eta_p^2 = 0.309$), indicating that control patients found significantly more targets per second ($M = 0.196$, $SD = 0.057$) than neglect patients ($M = 0.124$, $SD = 0.066$). The main effect of screen size, on the other hand, was not significant ($F(2,30) = 0.235$, $p = 0.792$), indicating that a comparable number of targets was found per second in all three screen sizes. The interaction was not significant either ($F(2,30) = 1.983$, $p = 0.155$). The
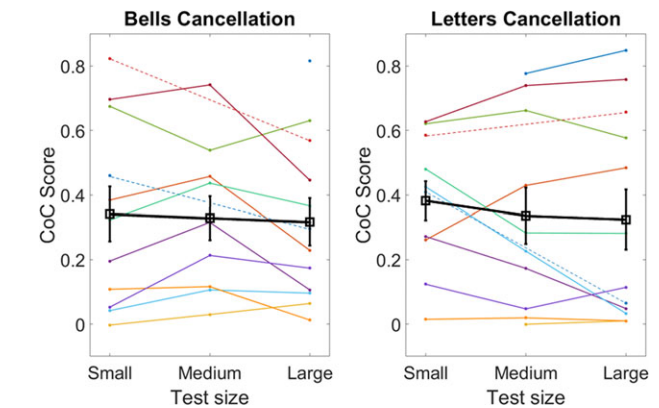

Figure 5. CoC scores over different digital test sizes. Neglect patients' performances in the three different touch screen formats of the digitalized Bells test and Letters cancellation test. Test scores of each patient were connected with a line (patients who failed to complete the medium test size version were indicated by a broken line). The bold lines represent mean values (including standard error) averaged over all patients.

same analysis applied on the Letters test also revealed a significant main effect of group ($F(1,13) = 8.624$, $p = 0.012$ $\eta_p^2 = 0.399$), indicating again that control patients on average found more targets per second ($M = 0.278$, $SD = 0.090$) than neglect patients ($M = 0.140$, $SD = 0.091$). The main effect of screen size was significant as well ($F(2,26) = 5.219$, $p = 0.012$); the interaction was not significant ($F(2,26) = 0.176$, $p = .839$). According to post hoc comparisons (Fisher's Least Significant Difference), more targets per second were found in condition TS large ($M = 0.235$, $SD = 0.120$) than in condition TS small ($M = 0.183$, $SD = 0.066$, $p < 0.05$, $d = 0.537$).

Search distance. Data are illustrated in Figure 6. Analysis of the Bells test revealed no main effect of screen size ($F(2,30) = 0.250$, $p = 0.781$) and interaction ($F(2,30) = 3.109$, $p = 0.059$), but a main effect of group ($F(1,15) = 7.357$, $p = 0.016$, $\eta_p^2 = 0.329$). Apparently search distance was smaller for control patients ($M = 0.550$, $SD = 0.093$) than for neglect patients ($M = 0.674$, $SD = 0.093$). For the Letters test, there was both a main effect of group ($F(1,13) = 5.486$, $p = 0.036$, $\eta_p^2 = 0.292$) and of screen size
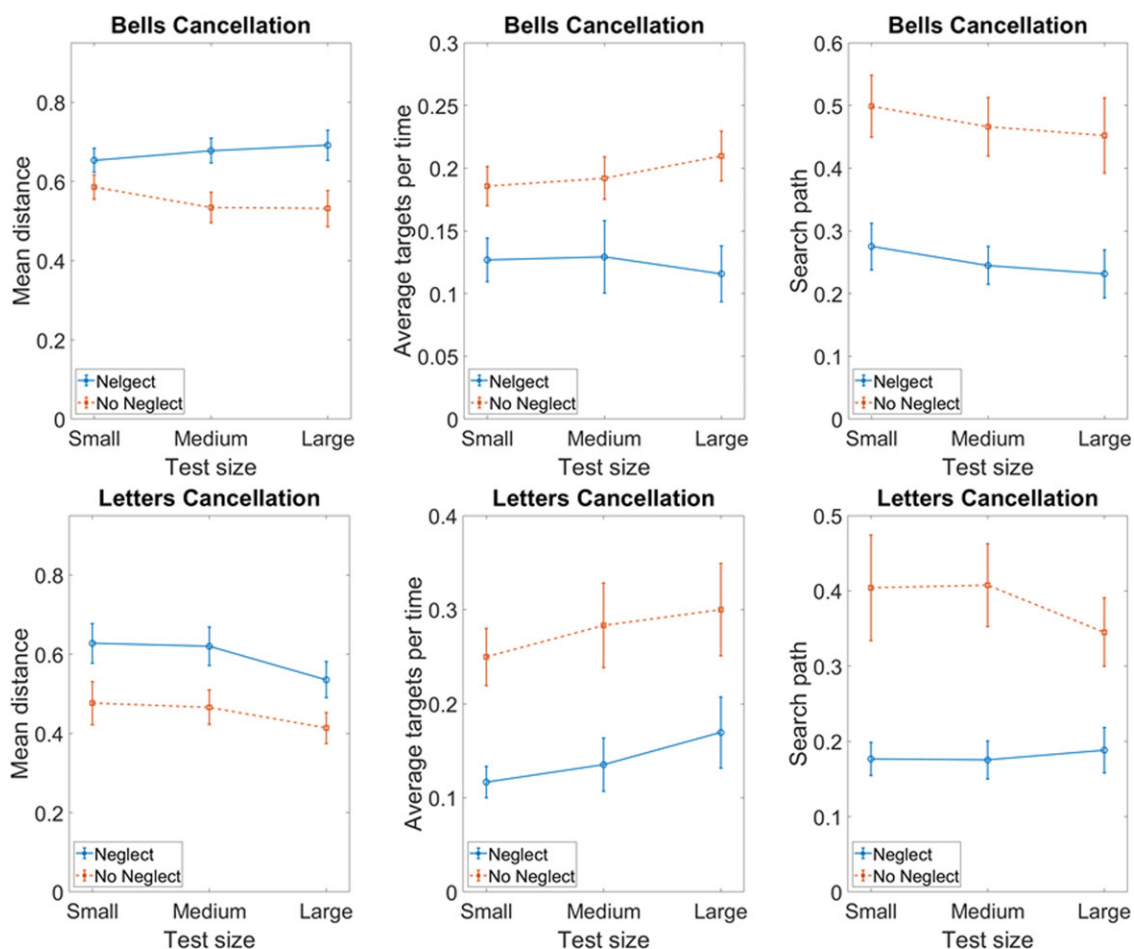
Figure 6. Additional search parameters over different test sizes. Averaged mean distance between to targets found in direct succession (left panel), averaged number of targets identified per time (middle panel), and averaged mean distance between two successive targets (right panel) (including standard error) in the three different touch screen formats of the digitalized cancellation tests in patients with and without spatial neglect (Neglect, No Neglect).

$(F(2,26) = 5.528, p = 0.010)$. The interaction was not significant $(F(2,26) = 0.244, p = 0.785)$. Again, search distance was smaller for control patients ($M = 0.452$, $SD = 0.116$) than for neglect patients ($M = 0.594$, $SD = 0.115$). Post hoc comparisons indicated that in the TS large version ($M = 0.474$, $SD = 0.120$) items found in direct succession were on average closer to one another than in the TS small ($M = 0.552$ $SD = 0.143$, $p = 0.026$, $d = 0.590$) and TS medium ($M = 0.543$, $SD = 0.127$, $p = 0.018$, $d = 0.557$) versions.

Search strategy. Data are illustrated in Figure 6. Analysis of this parameter for the Bells test did neither show a main effect of screen size ($F(2,30) = 2.503$, $p = 0.099$) nor an interaction ($F(2,30) = 0.002, p = 0.998$), while the main effect of group was significant $F(1,22) = 9.11$, $p < 0.01$, $\eta_p^2 = 0.502$. Control patients scored significantly higher ($M = 0.472$, $SD = 0.119$) than neglect patients ($M = 0.251$, $SD = 0.117$), indicating that search behavior of neglect patients was more strategic than the one of control patients. Results concerning the Letters test uncovered a main effect for group ($F(1,13) = 15.787$, $p = .002$ $\eta_p^2 = 0.548$), indicating that neglect patients search behavior was significantly more strategic ($M = 0.180$, $SD = 0.51$) than the one of right-hemispheric control patients ($M = 0.386$, $SD = 0.101$). There was neither a main effect of screen size ($F(2,26) = 0.513$, $p = 0.604$) nor an interaction ($F(2,26) = 0.1.177$, $p = 0.324$).

Prediction of spatial neglect by the additional parameters of search behavior

To determine if the three additional parameters of search behavior provided by the digital format can be used to differentiate between right-hemispheric patients with and without spatial neglect, SVM were used. First, the binary diagnosis (neglect vs. no neglect) was predicted separately for the Bells and the Letters tests across all screen sizes, using Leave-One-Subject-Out Cross-Validation. Results showed that this cross-participant classification across screen sizes was highly accurate for the Bells test and for the Letters test with average balanced accuracy scores of 97.92% and 88.19%, respectively. The training and test accuracies for all models are shown in Figure 7. Second, chi-square tests of independence indicated that the frequency of accurately predicted neglect and right-hemispheric control patients (see Table 2) was independent of the screen size for the Bells ($\chi^2(2) = 0.05$, $p = 0.973$) and Letters test ($\chi^2(2) = 0.18$, $p = 0.914$). To investigate if the machine learning models solely predict neglect diagnosis as a proxy for lesion size, we tested if the models could accurately differentiate if a participant had an above or below average lesion volume compared to the sample. Results showed low accuracy for models trained on both tests (Bells test 64.29%, Letter test 54.76%) indicating that predictions were largely made independent of lesion volumes.

Table 2. Frequency of correctly classified neglect and right-hemispheric control patients by screen size (TS small, TS medium, and TS large)

Screen size	Diagnosis	Prediction Accuracy	
		Bells cancelation test	Letters cancelation test
TS small	control (n = 8)	7	7
	neglect (n = 12)	12	9
TS medium	control (n = 8)	8	7
	neglect (n = 12)	12	12
TS large	control (n = 8)	8	7
	neglect (n = 12)	12	9
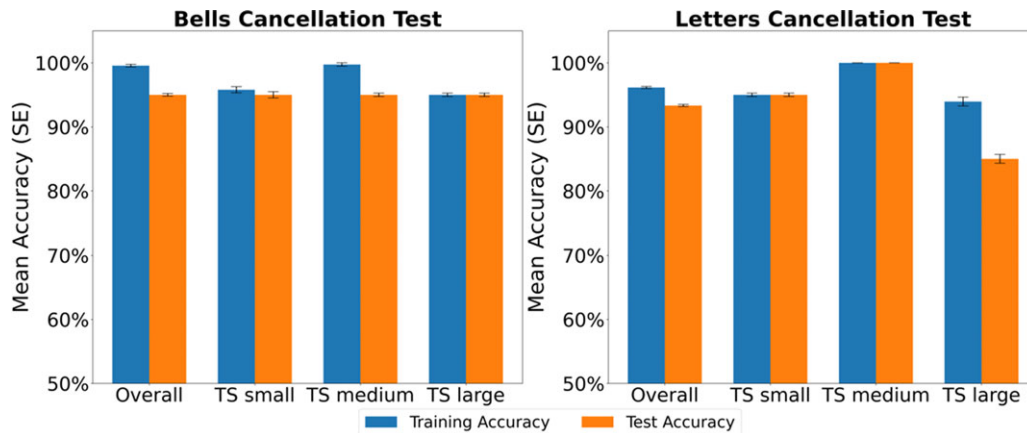
Figure 7. Machine learning model performance. Training and test performance of the machine learning models classifying the binary diagnosis "neglect vs. no neglect" in the right hemispheric patient sample overall and broken down by screen size (TS small, TS medium, and TS large).

Discussion

Paper-and-pencil versus digital test version

Several papers have acknowledged numerous perks of digitizing neuropsychological assessments in general (Bauer et al., 2012; Germine, Reinecke, & Chaytor, 2019) and neglect diagnostics specifically (Donnelly et al., 1999; Bonato, Priftis, Marenzi, Umiltà, & Zorzi, 2012; Bonato & Deouell, 2013). This has inspired the introduction of novel computer-based neglect assessments (Donnelly et al., 1999; Deouell, Sacher & Soroker, 2005; Bar-Haim, Kizony, Shahar, & Katz, 2006; List et al., 2008; Bonato, Priftis, Umiltà, & Zorzi, 2013; Dalmaijer et al., 2015; Villarreal et al., 2020). Digital versions have been argued to be more flexible, allowing to create several parallel versions of a specific task and therefore preventing learning effects from numerous repetitions of one identical version, for example, in the course of rehabilitation (Bonato & Deouell, 2013). They can further be created to be immediately adaptive to patients' individual performance (List et al., 2008). Moreover, digital formats could further increase a test's sensitivity by increasing the amount of information extractable from its data (Bonato & Deouell, 2013, Dalmaijer et al., 2015). However, previous studies also stressed the importance of validating digital formats (Bauer et al., 2012; Germine et al., 2019). The present paper is to our knowledge the first that systematically compared patients' performances between digital and analogous formats. Stroke patients' CoC derived from cancelation tasks seems robust to test digitization. Thus, it seems safe to introduce digitized diagnostic measures (at least in the scope of size variations as investigated in the present study) and keep the existing cut-off scores, without having to fear distortions in the CoC and related diagnostic decisions.

Test/display size of the cancelation tasks

Center of cancelation

Patients' CoCs did not seem to be impacted by test size either. Our analysis revealed that neglect patients ignored a comparable ratio of contralesional target stimuli, regardless of test size. This observation corresponds to previous findings on reference frames suggesting a dynamic view of the neglected area in space, depending on the respective behavioral goal of the subject. Karnath and Niemeier (2002) argued that the brain continuously organizes and re-organizes the representation of the same physical input according to the changing task requirements. The authors showed that whether or not neglect patients ignored certain spaces in a visual search task did not depend on the frame size itself, but rather on the relative location within the part of space they were asked to pay attention to. Patients were found to ignore the left half of space when asked to explore only that very segment but attended to it fully when it constituted the right half of a larger segment. Similar results were found by Baylis and colleagues (2004). The observation that removing targets once they are identified by patients reduces patient's attention frame and thus manages to draw patients' attention further into contralesional space (Mark et al., 1988; Keller, Volkening & Garbacenkaite, 2015) further supports this notion. In conclusion, these findings indicate that the neglect-specific egocentric bias seems to be robust to variations in screen size and provides a suitable explanation for the CoC's indifference to size changes observed in the present study. While our results based on a sample size of 12 continuously admitted neglect patients represent an initial estimate, further evidence based on larger samples is needed to endorse our findings.

Additional parameters of search behavior

In the additional digital parameters, neglect patients showed decreased search speed, increased search distance, and a more strategic search pattern than right-hemispheric control patients without neglect in both cancelation tasks.

Search speed and search distance. Deouell and colleagues (2005) had already shown that reaction times measured in a dynamic search task appear more sensitive than a common attention battery in illustrating neglect deficits and their recovery. Our finding that processing time of search behavior is also impaired is in agreement with these observations. Neglect patients are frequently observed to start working on tasks from the right side and effortfully drag their attention towards the contralesional hemispace. An eye-tracking study on reading behavior, for example, illustrated how straining it is for a neglect patient to advance further towards the neglected left. While healthy readers find the beginning of the next text line by performing long, pointed saccades, the investigated neglect patient moved leftward gradually (Karnath & Huber, 1992). Of course, the latter is much more time-consuming. Studies investigating the visual scanning and exploration behavior of neglect patients on photographs (Ptak, et al., 2009; Machner et al., 2012; Kaufmann et al., 2020) and videos (Machner et al., 2012) found that increased salience due to motion (Ptak et al., 2009) or contrasts (Machner et al., 2012) can help a patient attend to the neglected hemispace. Kaufmann and colleagues (2020) found neglect patients' perseverance to ipsilesional space under neutral conditions to be so distinct that it proved to be more sensitive in detecting neglect than common diagnostical measures. This exploration pattern might directly translate to our visual search tasks. Neglect patients may be more likely to move leftward inefficiently progressing from one stimulus to the next, coming across a target every now and then rather coincidentally. This process makes them more prone to miss a target if a distractor is closer to the current fixation and attracts patients' attention instead, resulting in a larger search distance.

Screen size-dependent performance was found in the Letters but not the Bells test, which could be caused by the different complexity of the tests. Neglect according to the Letter cancelation task is diagnosed if more than four contralesional stimuli are omitted, while a diagnosis based on the Bells test requires at least five omissions (Rorden & Karnath, 2010). Out of context, that doesn't seem a grave difference, however, the Letters test contains 60, Bells test only 35 targets hidden among distractors. Relatively speaking, the cut-offs represent 6 % versus 14 % omissions, indicating that healthy individuals are more likely to omit bells than "A"s. Automatized letter recognition is known to be superior to object identification (Denckla & Rudel, 1974), which might explain why in the Letters but not the Bells test targets were identified faster in the large screen size than the small one. Since automatized reading depends on how well letters are recognizable, enlarging letters in our paradigm likely improved participants' perception and search efficiency by facilitating target/distractor discrimination. Bells' more effortful shape-identification might not benefit as much from a larger depiction. Patients searched significantly faster only between the small and the large variant of the Letters test. The search was more thorough in the large version compared to the small and medium when normalized for screen size. Since the size difference between medium and large (59 %) is greater than between small and medium (28 %), 59 % magnification seems sufficient to decrease the likelihood of missing closeby targets (thus decreasing search distance), while improving search speed requires a larger increase in test size.

Search strategy. In contrast to the fairly straightforward measures for search speed and search distance, it is rather hard to come up with a universal indicator of search strategy, since a strategic search can be performed in many different ways. The measure we defined as "search strategy" in the present study should cover strategies typically applied by healthy individuals (cf. Figure 2; Warren, Moore, & Vogtle, 2008). Interestingly, we found that neglect patients' search behavior was more "strategic" (according to our definition) than that of right-hemispheric control patients in both Bells and Letters test. While neglect patients more frequently applied strategies such as those shown in Figure 2, right-hemispheric controls either searched in a less "strategic" manner or applied a strategy different from the ones typically applied by healthy individuals (Warren et al., 2008). While this finding might seem surprising at first glance, previous research provided evidence that neglect patients do not generally exhibit impairments in search strategy (Donnelly et al., 1999; Mark et al., 2014; Ten Brink et al., 2018). Donnelly and colleagues (1999) generated 16 different strategic search patterns and investigated which ones were applied by healthy participants as well as by right-hemispheric patients with and without neglect. Although neglect patients tended to apply different search strategies than the majority of healthy participants and non-neglect patients, their pattern matched some of the authors' predefined strategic search paths. More specifically, neglect patients most frequently applied a strategy that mirrored the most common strategy used by healthy controls. The favored strategy reported by Donelly and colleagues was equivalent to the one illustrated in Figure 2 C. While controls started from the top left corner and worked their way to the right, neglect patients started from the top right corner and proceeded leftward. In accordance with Donelly, our neglect patients started 100% from the right side in the bell test and 94% in the letter test. However, since our measure accounted for strategies starting from both sides (left or right), both directions were considered equally strategic. Our results indicate that neglect patients seem to follow a (potentially predefined) line- or row-wise search pattern, while patients without neglect rather turn to overall close items. Neglect patients' previously mentioned search efforts might make them more susceptible to applying search strategies, potentially to compensate for their lacking overall search efficiency which might impair their detection of targets in the close proximity of the last hit.

With regards to the diagnostic use of the three process measures (i.e., search speed, search distance, and search strategy), machine learning classifications indicated that these variables can be used to differentiate neglect patients from right-hemispheric controls reliably, using parsimonious modeling approaches. Particularly for the Letters test, the overall small differences between training and test accuracy for these cross-participant classifications (see Figure 7) indicate that the models generalize well, which is crucial for potential applications of such measures. For instance, digital tests could capture these measures in real-time and predict the diagnosis already at the early stages of the testing procedure, which would be beneficial if, for example, the test has to be aborted. This, in turn, could serve as a basis for adaptive and time-saving diagnostic procedures that are less strenuous for patients. While overall predictions were highly accurate in both tests (97.92% for the Letters test and 88.19% for the bells cancelation test), it is important to note that predictions for the Letters test predictions were less accurate and showed a larger variation between training and test samples than for the bells cancelation task. Here, future research with larger patient samples is

needed to further evaluate such differences in the diagnostic value of process measures between tests. With regards to screen size, our analyses showed that model predictions for both tests were independent of screen size. For further studies and potential practical applications, this indicates that additional measures obtained from digitized tests can be used to reliably classify neglect regardless of screen size. Nonetheless, future research with larger sample sizes is required to confirm the robustness of our models.

Conclusion

The present results allow an optimistic outlook on the digitization of cancelation tasks. Changes in test format (paper-and-pencil vs. digital) and in screen size do not seem to bias patients' CoC measure, which often serves for diagnostic decision-making in spatial neglect. This robustness opens the possibility to optimize some visual parameters for more efficient testing. Increasing the stimulus size in the Letters test seems to help patients identify targets more quickly, which would make diagnosis less time-consuming for the examiner and less exhausting for the patients. Machine learning methods indicated that new search parameters derived from computerized tests could help differentiate neglect and non-neglect patients. The latter is an interesting new perspective because in clinical practice it is often not possible to perform cancelation tasks to the end (which is mandatory to calculate the CoC measure). If neglect-specific performance features can be extracted from variables such as search speed, search distance, and search strategy neglect diagnosis might become possible even if the test is discontinued. Future studies are needed to investigate the latter.

References

Anderson, B. (1997). Pieces of the true crossover effect in neglect. *Neurology*, *49*, 809–812. https://doi.org/10.1212/wnl.49.3.809

Bar-Haim Erez, A., Kizony, R., Shahar, M., & Katz, N. (2006).Visual spatial search task: A computerized assessment and training program. *Published Proceedings of the ICDVRAT conference*. Esberg, Denmark, September.

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: Joint position paper of the American academy of clinical neuropsychology and the national academy of neuropsychology. *Archives of Clinical Neuropsychology*, *27*, 362–373. https://doi.org/10.1093/arclin/acs027

Baylis, G. C., Baylis, L. L., & Gore, C. L. (2004). Visual neglect can be object-based or scene-based depending on task representation. *Cortex*, *40*, 237–246. https://doi.org/10.1016/S0010-9452(08)70119-9

Becker, E., & Karnath, H. O. (2007). Incidence of visual extinction after left versus right hemispheric stroke. *Stroke*, *38*, 3172–3174

Bonato, M., & Deouell, L. Y. (2013). Hemispatial neglect: Computer-based testing allows more sensitive quantification of attentional disorders and recovery and might lead to better evaluation of rehabilitation. *Frontiers in Human Neuroscience*, *7*, 1–4. https://doi.org/10.3389/fnhum.2013.00162

Bonato, M., Priftis, K., Marenzi, R., Umiltà, C., & Zorzi, M. (2012). Deficits of contralesional awareness: A case study on what paper-and-pencil tests neglect. *Neuropsychology*, *26*, 20–36. https://doi.org/10.1037/a0025306

Bonato, M., Priftis, K., Umiltà, C., & Zorzi, M. (2013). Computer-based attention-demanding testing unveils severe neglect in apparently intact patients}. *Behavioural Neurology*, *26*, 179–181. https://doi.org/10.3233/BEN-2012-129005

Bowers, D., & Heilman, K. M. (1980). Pseudoneglect: Effects of hemispace on a tactile line bisection task. *Neuropsychologia*, *18*, 491–498. https://doi.org/10.1016/0028-3932(80)90151-7

Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. Proceedings - *20th International Conference on Pattern Recognition (ICPR)*, 3121–3124. https://doi.org/10.1109/ICPR.2010.764

Chen, S., Stromer, D., Alabdalrahim, H. A., Schwab, S., Weih, M., & Maier, A. (2020). Automatic dementia screening and scoring by applying deep learning on clock-drawing tests. *Scientific Reports*, *10*, 1–11.

Corbetta, M., & Shulman, G. L. (2011). Spatial neglect and attention networks. *Annual Review of Neuroscience*, *34*. https://doi.org/10.1146/annurev-neuro-061010-113731

Dalmaijer, E. S., Van der Stigchel, S., Nijboer, T. C. W., Cornelissen, T. H. W., & Husain, M. (2015). CancellationTools: All-in-one software for administration and analysis of cancellation tasks. *Behavior Research Methods*, *47*, 1065–1075. https://doi.org/10.3758/s13428-014-0522-7

De Haan, B., Clas, P., Juenger, H., Wilke, M., & Karnath, H. O. (2015). Fast semi-automated lesion demarcation in stroke. *NeuroImage: Clinical*, *9*, 69–74. https://doi.org/10.1016/j.nicl.2015.06.013

Denckla, M. B., & Rudel, R. (1974). Rapid "Automatized" naming of pictured objects, colors, letters and numbers by normal children. *Cortex*, *10*, 186–202. https://doi.org/10.1016/S0010-9452(74)80009-2

Deouell, L. Y., Sacher, Y., & Soroker, N. (2005). Assessment of spatial attention after brain damage with a dynamic reaction time test. *Journal of the International Neuropsychological Society*, *11*, 697–707. doi: 10.1017/S1355617705050824

Donnelly, N., Guest, R., Fairhurst, M., Potter, J., Deighton, A., & Patel, M. (1999). Developing algorithms to enhance the sensitivity of cancellation tests of visuospatial neglect. *Behavior Research Methods, Instruments, and Computers*, *31*, 668–673. https://doi.org/10.3758/BF03200743

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, *14*, 91–118.

Ferber, S., & Karnath, H. (2001). How to assess spatial neglect: Line bisection or cancellation tasks ? *Journal of Clinical and Experimental Neuropsychology*, *23*, 599–607.

Fullerton, K. J., Stout, R. W., & McSherry, D. (1986). Albert's test: A neglected test of perceptual neglect. *The Lancet*, *327*, 430–432. https://doi.org/https://doi.org/10.1016/S0140-6736(86)92381-0

Gauthier, L., Dehaut, F., & Joanette, Y. (1989). The bells test: A quantitative and qualitative test for visual neglect. *International Journal of Clinical Neuropsychology*, *11*, 49–54.

Germine, L., Reinecke, K., & Chaytor, N. S. (2019). Digital neuropsychology: Challenges and opportunities at the intersection of science and software. *Clinical Neuropsychologist*, *33*, 271–286. https://doi.org/10.1080/13854046.2018.1535662

Johannsen L., & Karnath H.-O. (2004). How efficient is a simple copying task to diagnose spatial neglect in its chronic phase? *Journal of Clinical and Experimental Neuropsychology*, *26*, 251–256. https://doi:10.1076/jcen.26.2.251.28085

Karnath, H. O., & Huber, W. (1992). Abnormal eye movement behaviour during text reading in neglect syndrome: A case study. *Neuropsychologia*, *30*, 593–598. https://doi.org/10.1016/0028-3932(92)90062-Q

Karnath, H. O., Mandler, A., & Clavagnier, S. (2011). Object-based neglect varies with egocentric position. *Journal of Cognitive Neuroscience*, *23*, 2983–2993. https://doi.org/10.1162/jocn_a_00005

Karnath, H. O., & Rorden, C. (2012). The anatomy of spatial neglect. *Neuropsychologia*, *50*, 1010–1017. https://doi.org/10.1016/j.neuropsychologia.2011.06.027

Karnath, H.-O., & Niemeier, M. (2002). Task-dependent differences in the exploratory behaviour of patients with spatial neglect. *Neuropsychologia*, *40*, 1577–1585.

Kaufmann, B. C., Cazzoli, D., Pflugshaupt, T., Bohlhalter, S., Vanbellingen, T., Müri, R. M., Nef, T., & Nyffeler, T. (2020). Eyetracking during free visual exploration detects neglect more reliably than paper-pencil tests. *Cortex*, *129*, 223–235. https://doi.org/10.1016/j.cortex.2020.04.021

Keller, I., Volkening, K., & Garbacenkaite, R. (2015). Stimulus-dependent modulation of visual neglect in a touch-screen cancellation task. *Neuropsychology*, 29, 417–420. https://doi.org/10.1037/neu0000146

Li, D., Karnath, H.-O., & Rorden, C. (2014). Egocentric representations of space co-exist with allocentric representations: evidence from spatial neglect. *Cortex*, 58, 161–169. https://doi.org/10.1016/j.cortex.2014.06.012.

List, A., Brooks, J. L., Esterman, M., Flevaris, A. V., Landau, A. N., Bowman, G., Stanton, V., Vanvleet, T. M., Robertson, L. C., & Schendel, K. (2008). Visual hemispatial neglect, re-assessed. *Journal of the International Neuropsychological Society*, 14, 243–256. doi: 10.1017/S1355617708080284

Machner, B., Dorr, M., Sprenger, A., von der Gablentz, J., Heide, W., Barth, E., & Helmchen, C. (2012). Impact of dynamic bottom-up features and top-down control on the visual exploration of moving real-world scenes in hemispatial neglect. *Neuropsychologia*, 50, 2415–2425. https://doi.org/10.1016/j.neuropsychologia.2012.06.012

Mark, V. M., Woods, A. J., Ball, K. K., Roth, D. L., & Mennenmeier, M. (2014). Disorganized search on cancellation is not a consequence of neglect. *Neurology*, 63, 78–84.

Mark, V. W., Kooistra, C. A., & Heilman, K. M. (1988). Hemispatial neglect affected by non-neglected stimuli. *Neurology*, 38, 1207–1211. https://doi.org/10.1212/wnl.38.8.1207

McCourt M.E., & Jewell G. (1999). Visuospatial attention in line bisection: Stimulusmodulation of pseudoneglect. *Neuropsychologia* 37, 843–855. https://doi.org/10.1016/S0028-3932(98)00140-7

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., & Louppe, G. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Ptak, R., Golay, L., Müri, R. M., & Schnider, A. (2009). Looking left with left neglect: The role of spatial attention when active vision selects local image features for fixation. *Cortex*, 45, 1156–1166. https://doi.org/10.1016/j.cortex.2008.10.001

Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., & Karnath, H. O. (2012). Age-specific CT and MRI templates for spatial normalization. *NeuroImage*, 61, 957–965. https://doi.org/10.1016/j.neuroimage.2012.03.020

Rorden, C., Hjaltason, H., Fillmore, P., Fridriksson, J., Kjartansson, O., Magnusdottir, S., & Karnath, H. O. (2012). Allocentric neglect strongly associated with egocentric neglect. *Neuropsychologia*, 50, 1151–1157. https://doi.org/10.1016/j.neuropsychologia.2012.03.031

Rorden, C., & Karnath, H.-O. (2010). A simple measure of neglect severity. *Neuropsychologia*, 48, 2758–2763. https://doi.org/10.1016/j.neuropsychologia.2010.04.018.

Schendel, K. L., & Robertson, L. C. (2002). Using reaction time to assess patients with unilateral neglect and extinction. *Journal of Clinical and Experimental Neuropsychology*, 24, 941–950. https://doi.org/10.1076/jcen.24.7.941.8390

Souillard-Mandar, W., Penney, D., Schaible, B., Pascual-Leone, A., Au, R., & Davis, R. (2021). DCTclock: clinically-interpretable and automated artificial intelligence analysis of drawing behavior for capturing cognition. *Frontiers in Digital Health*, 3, 1–10.

Ten Brink, A. F., Visser-Meily, J. M. A., & Nijboer, T. C. W. (2018). What does it take to search organized? the cognitive correlates of search organization during cancellation after stroke. *Journal of the International Neuropsychological Society*, 24, 424–436. https://doi.org/10.1017/S1355617717001254

Villarreal, S., Linnavuo, M., Sepponen, R., Vuori, O., Jokinen, H., & Hietanen, M. (2020). Dual-task in large perceptual space reveals subclinical hemispatial neglect. *Journal of the International Neuropsychological Society*, 26, 993–1005. https://doi.org/10.1017/S1355617720000508

Warren, M., Moore, J. M., & Vogtle, L. K. (2008). Search performance of healthy adults on cancellation tests. *American Journal of Occupational Therapy*, 62, 588–594. https://doi.org/10.5014/ajot.62.5.588

Weintraub, S., & Mesulam, M.-M. (1985). Mental state assessment of young and elderly adults in behavioral neurology. In M.-M. Mesulam (Ed.), *Principles of behavioral neurology* (pp. 71–123). Philadelphia, PA: F.A. Davis Company.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine*, 30, 377–399.