



Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex



Research report

Selection and application of familiar and novel tools in patients with left and right hemispheric stroke: Psychometrics and normative data

Ilka Buchmann ^{a,b} and Jennifer Randerath ^{a,b,*}

^a University of Konstanz, Konstanz, Germany

^b Lurija Institute for Rehabilitation and Health Sciences at the University of Konstanz, Schmieder Foundation for Sciences and Research, Allensbach, Germany

ARTICLE INFO

Article history:

Received 21 January 2017

Reviewed 26 February 2017

Revised 27 May 2017

Accepted 1 June 2017

Action editor Georg Goldenberg

Published online 23 June 2017

Keywords:

Limb apraxia

Tool-use

Mechanical problem solving

Psychometric properties

DILA-S

ABSTRACT

Frequently left brain damage (LBD) leads to limb apraxia, a disorder that can affect tool-use. Despite its impact on daily life, classical tests examining the pantomime of tool-use and imitation of gestures are seldom applied in clinical practice. The study's aim was to present a diagnostic approach which appears more strongly related to actions in daily life in order to sensitize applicants and patients about the relevance of the disorder before patients are discharged.

Two tests were introduced that evaluate actual tool selection and tool-object-application: the Novel Tools (NTT) and the Familiar Tools (FTT) Test (parts of the DILA-S: Diagnostic Instrument for Limb Apraxia – Short Version). Normative data in healthy subjects ($N = 82$) was collected. Then the tests were applied in stroke patients with unilateral left brain damage (LBD: $N = 33$), a control right brain damage group (RBD: $N = 20$) as well as healthy age and gender matched controls (CL: $N = 28$, and CR, $N = 18$).

The tests showed appropriate interrater-reliability and internal consistency as well as concurrent and divergent validity. To examine criterion validity based on the well-known left lateralization of limb apraxia, group comparisons were run. As expected, the LBD group demonstrated a high prevalence of tool-use apraxia (NTT: 36.4%, FTT: 48.5%) ranging from mild to severe impairment and scored worse than their control group (CL). A few RBD patients did demonstrate impairments in tool-use (NTT: 15%, FTT: 15%). On a group level they did not differ from their healthy controls (CR). Further, it was demonstrated that the selection and application of familiar and novel tools can be impaired selectively.

Our study results suggest that real tool-use tests evaluating tool selection and tool application should be considered for standard diagnosis of limb apraxia in left as well as right brain damaged patients.

© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author. University of Konstanz, Konstanz, Germany.

E-mail address: J_Randerath@hotmail.com (J. Randerath).

<http://dx.doi.org/10.1016/j.cortex.2017.06.001>

0010-9452/© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Abbreviations

AST	apraxia screen of TULIA
CL	control group of healthy participants matched for left brain damaged patients in age, sex, education and tested hand (left)
CR	control group of healthy participants matched for right brain damaged patients in age, sex, education and tested hand (right)
FTT	Familiar Tools Test
HC	healthy control participants
KAS	Kölner Apraxie Screening
LBD	left brain damage due to stroke
NTT	Novel Tools Test
RBD	right brain damage due to stroke
SAST	short apraxia screening test
STIMA	short test for ideomotor apraxia
TULIA	test of upper limb apraxia

1. Introduction

The term apraxia refers to ‘higher level’ disorders of motor control (Goldenberg, 2013a). It occurs independent of pure motoric deficits like hemiplegia or cognitive deficits like aphasia. Goldenberg (2013c) describes apraxia as a motor-cognitive disorder with impairments typically in imitation, execution of communicative gestures and/or usage of tools and objects, respectively. Apraxia in most cases follows left brain damage (LBD), typically affects both sides of the body, and often co-occurs with aphasia (Donkervoort, Dekker, Van Den Ende, & Stehmann-Saris, 2000; Goldenberg, 2011). Besides stroke or traumatic brain injury lesions, it can also occur with neurodegenerative lesions such as dementia (e.g., Chandra, Isaac, & Abbas, 2015; Cotelli, Manenti, Brambilla, & Balconi, 2014) or corticobasal syndrome (e.g., Burrell, Hornberger, Vucic, Kiernan, & Hodges, 2014; Stamenova et al., 2015). Traditionally, it is tested by asking the patient to pantomime object use movements (e.g., “Show me how to use a toothbrush.”) or by imitation of gestures. However, patients can also be impaired in the actual application of real tools and objects (tool-use apraxia; Goldenberg, 2011). Patients may choose the wrong objects (e.g., soap instead of toothpaste for brushing the teeth) and/or perform the movement wrongly or not at all (e.g., not moving the fork into the mouth or not even exploring the fork).

It is a known phenomenon that the classic tasks of imitation and pantomime are more sensitive to erroneous behavior compared to real tool-use tasks (De Renzi, Faglioni, & Sorgato, 1982; Randerath, Goldenberg, Spijkers, Li, & Hermsdoerfer, 2011). The differences between performance in for example the pantomime task on the one hand and real tool-use tasks on the other hand could either be explained by different domains of limb apraxia or by differing workload, respectively (Belanger, Duffy, & Coelho, 1996). The “disconnection hypothesis” proposes that the described tasks are solved by differing independent neuropsychological processes. In

contrast, the “severity hypothesis” implies that the different tasks are solved by a common mechanism and that there is a continuum of task-difficulty. Previous studies demonstrated that improved behavioral performance when actually applying the tool compared to pantomiming was especially found in more complex actions (Hermsdoerfer, Li, Randerath, Roby-Brami, & Goldenberg, 2013; Randerath et al., 2011). The greater extent of contextual information occurred to curtail action opportunities (affordances) provided in real tool-use tasks which might minimize demands on working memory involved in the retrieval and integration of information necessary for planning the action. However, despite the relative improvement of behavior in real tool-use settings, patients still can be affected significantly (Randerath et al., 2011).

Obviously, limb apraxia has high relevance for maintaining independence in daily life activities (Goldenberg & Hagmann, 1998b; Goldenberg, Daumüller, & Hagmann, 2001; Unsal-Delialioglu, Kurt, Kaya, Culha, & Ozel, 2008). Patients who have been classified as apraxic by pantomime or imitation tests are more often dependent on nursing staff (Poeck, 2006; Wu, Burgard, & Radel, 2014) and return less often to work than non-apraxic patients (Dovern, Fink, & Weiss, 2011; Wang, Kapellusch, & Garg, 2014). Additionally, the severity of apraxia predicts the rehabilitation success for patients with hemiplegia (Dovern et al., 2011; Hanna-Pladdy, Heilman, & Foundas, 2003).

Despite the fact that the use of tools and objects is especially relevant to empower patients for an independent daily live (De Renzi & Lucchelli, 1988; Goldenberg & Hagmann, 1998a; Mayer, Reed, Schwartz, Montgomery, & Palmer, 1990; Poeck, 1983), limb apraxia assessments still are not conducted for standard diagnostics in most clinics, notwithstanding that validated test batteries are available that typically include the traditional tests of imitating gestures and pantomime of tool-use, e.g., the TULIA (test of upper limb apraxia; Vanbellingen et al., 2010), its short form, the AST (apraxia screen of TULIA; Vanbellingen, 2012, 2013), the KAS (Kölner Apraxie Screening; Weiss, Kalbe, Kessler, & Fink, 2013) or the STIMA (short test for ideomotor apraxia; Tessari, Toraldo, Lunardelli, Zadini, & Rumiati, 2015).

Several points may explain why this significant disorder still does not receive enough attention to be diagnosed. First, the impact of limb apraxia may be overlooked, because of other impairments receiving amplified attention by caregivers and patients. For example, hemiplegia and aphasia are two obvious detectable co-occurring deficits that patients and relatives are concerned about and want to be taken care of. Further, the traditional tests included in standard batteries (imitation or pantomime tasks) may appear quite abstract to the patient and the therapist. Remediation could be achieved by introducing tests with real tools and objects that directly relate to activities of daily living. Thus far, there are only few tests which directly assess tool-use apraxia with real objects. Even the newly developed test batteries that include real tool-use have a limited amount of items devoting to the impairment. For example, the SAST (short apraxia screening test; Leiguarda, Clarens, Amengual, Drucaroff, & Hallett, 2014) only has two items included: to rotate a coin between the fingers and to use a nail cutter, whereby the first task is rather a

mechanical motor task testing fine motor skills instead of the application of tools. Further, the KAS reports that preliminary versions included real tool-use and complex tasks but then were removed for the final version due to missing sensitivity and high material effort (Weiss et al., 2013).

Thus, there are few approaches available to diagnose difficulties in the actual use of novel or familiar tools in patients. Further, of these few approaches most concentrate on difficulties with tool application but neglect to look at problems with tool selection, which undoubtable is highly relevant in daily activities. We propose that real object use has to be included into the standard assessment of patients with brain damage in order to obtain an impression about the effects of apraxia on daily life activities and thereby to sensitize patients, relatives as well as clinicians about the relevance of the disorder.

Ideally, both familiar tool-use and novel tool-use should be tested in order to better locate the source of a patient's difficulties. The use of novel tools and objects for example presupposes familiarity with general principles of physics and mechanics rather than reliance on object semantics (Hegarty, 2004; Osiurak, Jarry, Lesourd, Baumard, & Le Gall, 2013; Povinelli, 2000; Zago & Lacquaniti, 2005). In contrast, the FTT is assumed to rely on both, on the retrieval of functional knowledge and object semantics from memory as well as on mechanical problem solving (Goldenberg & Spatt, 2009; Hodges, Spatt, & Patterson, 1999; Randerath et al., 2011).

For this purpose we developed a Familiar Tools Test (FTT) and adapted the Novel Tools Test (NTT) from Goldenberg and Hagmann (1998b) for the here used DILA-S (Diagnostic Instrument for Limb Apraxia – Short Version). Both tool-use tests include actual tools and objects to manipulate. In order to provide a suitable diagnostic approach for clinical settings, care was taken to develop a procedure that is suitable for patients with brain damage including those with only minimal to moderate comprehension abilities.

The here introduced tool-use tasks allow the separate evaluation of performance in tool selection and application. The evaluation of tool selection is based on whether patients need more than one attempt to select the best suitable tool out of three to manipulate the recipient object. The subsequent tool application is separated into a score that assesses trial and error attempts for executing the correct action (Execution Scale) as well as a score that assesses qualitatively how the action is produced (4-point Production Score). For each item four parameters are evaluated: grip-formation, grip-orientation, movement-content and movement-orientation.

The study by Goldenberg and Spatt (2009) used a real tool-use approach that is similar to ours, but there are some important differences. First, their normative data was based on a group of only 15 controls. Second, the novel tool-use test and the familiar tool-use test were less controlled with respect to task differences. Whereas the current FTT follows the same principle as the NTT (selecting the correct tool out of three to apply to the recipient object), the familiar tools task in the Goldenberg and Spatt (2009) study included the presentation of only one tool and the participant had to select the correct recipient out of five objects on a rack positioned behind the presented tool. In addition, a qualitative assessment of the action performance was not provided. We here included a

qualitative 4-point Production Score because error-profiles of individual patients may reveal helpful information for further diagnostics and therapeutic approaches (Deutsch-Lezak, Howieson, Bigler, & Tranel, 2012).

Normative data based on 82 healthy participants was collected. Interrater reliability, test consistency, concurrent, divergent and criterion validity were evaluated. Communicating with clinical personnel over the past years we received the impression that the classical tests of imitation and pantomime appear rather abstract to therapists, patients and their relatives. Low test-acceptance may be a strong reason for why apraxia diagnostics are not taken into account in the clinical routine. In order to receive an estimate for acceptance of the DILA-S we additionally adapted the questionnaire AKZEPT-L (Kersting, 2008) and collected feedback from local occupational therapists, who applied the tests.

2. Methods

2.1. Participants

To obtain normative data 82 healthy volunteers were tested. They were aged between 20 and 79 years, 63.4% were female and all were right handed (diagnosed with lateralization quotient ≥ 60 ; Salmasso & Longoni, 1983). Because patients may be forced to use their non-dominant left hand due to hemiparesis, half of the participants were tested with their right hand and half of the participants used their left hand.

Patients were recruited from the neurorehabilitation clinic “Kliniken Schmieder” in Allensbach, Germany. The patients did not require intensive care, were able to participate actively in therapy sessions and were resilient during 30 min of therapy. Patients were excluded from the study if they had any neurological or psychiatric disorder or did not speak German before stroke onset. Further, patients were not included in the study if they did not understand the instructions of the “Token Test” which tests speech comprehension. Patients were recruited based on the medical record. In order to post-hoc verify unilaterality, patients' MRI or CT scans were retrieved from their clinical records in the rehabilitation facility or were sent by the admission clinic. Patients were excluded from further analyses if it was not possible to obtain the MRI or CT scan. A total of 53 right handed stroke patients (diagnosed with lateralization quotient ≥ 60 ; Salmasso & Longoni, 1983) were included in the current study.

Of these, 33 patients had left hemisphere lesions (LBD). They were in the subacute (stroke onset 3 weeks – six months ago; 93.9%) or chronic phase (stroke onset more than six months ago; 6.1%). The patients were aged between 30 and 79 years, 48.5% were female. 20 patients had right hemisphere damage (RBD) and were included as a patient control group. They all were in the subacute phase. The patients were aged between 27 and 78 years, 55% were female.

All patients were tested with their ipsilesional hand and had no general difficulties to comply with our task instructions.

For group comparisons, healthy controls (HC) were matched to the recruited LBD (CL) and RBD (CR) patients for age and gender. The CL group included 28 HC using their left

hand for the requested activities. The CL and LBD groups did not differ in age ($p = .402$), sex ($p = .609$) and education ($p = .289$). The CR group, who used the right hand, contained 18 HC. CR and RBD groups also did not differ in age ($p = .823$), sex ($p = .752$) and education ($p = .164$). Please see [Table 1](#) for demographic information.

The study design was approved by the ethical committee of the University of Konstanz. All healthy participants and patients were taking part in the study voluntarily. Informed consent was obtained and privacy rights were observed. The study was conducted in accordance with the Declaration of Helsinki.

2.2. Developed assessment of tool-use apraxia

Two tests were developed, a NTT and a FTT. The Novel Tools Test was based on the one by [Goldenberg and Hagmann \(1998b\)](#). Originally, five cylinders of the Goldenberg & Hagmann set and six new ones were used. In each trial three tools and one wooden cylinder in a socket were presented in front of the participants (for example see [Fig. 1A](#)). The participants had to choose the tool that is best suitable to lift the cylinder out of the socket without dropping it. Inspired by the NTT, we developed the FTT. The participants were required to choose the correct tool to handle the recipient object and then apply it (see for example [Fig. 1B](#)).

The original version of the NTT and FTT each consisted of eleven different items (10 test items and 1 practice item). For each item the object to be manipulated was presented centrally in front of the participant on the table and three tools lying next to each other were positioned in front of the object. Whereas the tools could occur in different sets, the object to be manipulated only was presented once per test. Per item there was only one corresponding tool. To avoid any influences of hemiplegia in the FTT or NTT, all tasks could be managed single-handedly. While the original version of the NTT and FTT took up to 20 min per test, the below described short versions take approximately half of the time and are better suitable to fit into the daily routine in clinical practice. The short versions were created by selecting the five most sensitive items out of the ten original test items per test. The item-analysis was run on a data-set of 36 patients whose major damage was in the left hemisphere. Sensitivity was determined by using the discrimination score P ($P = \Sigma$ points of patients/ Σ max. possible points) and the selectivity of each item (correlation of each item with the whole subtest). Further, in order to facilitate and confirm comprehension also in aphasic patients, three items were suggested for practice trials. In order to help the examiner with the correct

interpretation of task comprehension, and in order to facilitate the start of the task for the patients, these practice items were chosen based on the idea that the probability of solving them (at least partly) is high. The selected practice items were solved correctly by more than 90.0% of the 36 patients. We recommend making use of at least one practice item to confirm task comprehension. If the patient immediately understands the task, the test items ($n = 5$) can be presented. Please see the evaluation sheets in the supplementary material that include images of the used objects per setting as well as the below described evaluation criteria. Further, the sheets inform about whether a patient's performance should be considered as apraxic.

For the evaluation of the NTT and FTT, three scales were used: one to evaluate tool selection and two to assess the suitability of tool application. For the selection of the correct tool it was evaluated how many attempts a participant needed to choose the correct tool. A tool was determined as being selected as soon as the participants made an effort to apply it at the recipient object (as opposed to just lifting and looking at the tool). If the first attempt was correct, then 2 points (*first correct*) were given, at second attempt 1 point (*second correct*) or else 0 points were noted (e.g., *total error*: the person did not detect the correct tool or tried the other two tools before using the correct one). For the application of the tools the so-called Execution Scale was similarly evaluated with 2 points (*first correct*) for showing immediately the correct use, 1 point (*second correct*) for trying movements leading to suitable tool-use and 0 points (*total error*) when the patient needed more than two attempts before using it correctly, or when the patient omitted the action because he or she had no idea of how to use the tool together with the object. In order to decide whether an action has been executed correctly, test evaluation sheets were used that describe clearly defined action characteristics for each item. These action characteristics had to be fulfilled in order for the action to be classified as correct. The description of each action's characteristics includes information about four parameters: grip-formation, grip-orientation, movement-content and movement-orientation. In order to provide a measure that informs qualitatively about the type of difficulties patients have with the task, the 4-point Production Score was included that evaluated each of the parameters (see supplementary material "Evaluation sheets" for production criteria). The participant received each one point for the grip-formation, the grip-orientation (thumb-direction on tool-handle), the movement-content and the movement-orientation for each item. The 4-point Production Score was included in order to provide potentially essential information for therapists about what aspects to primarily train with the

Table 1 – Demographic and clinic data.

Group	LBD $n = 33$	CL $n = 28$	RBD $n = 20$	CR $n = 18$
Gender: male/female	17/16	12/16	9/11	7/11
Age	60.45 (30–79)	57.64 (25–79)	59.00 (27–78)	59.44 (24–78)
days since stroke onset	98.12 (21–784)	n. a.	56.05 (23–102)	n. a.
Aphasia: No	11	28	14	18
Broca (mild/moderate/severe)	22 (6/5/11)		6 (6/0/0)	
Wernicke (mild/moderate/severe)	20 (9/3/8)		1 (1/0/0)	
(n.a.: not applicable).				



Fig. 1 – (A) Novel Tools Test. The one tool has to be selected that is best suitable to lift the cylinder out of the socket in a safe manner. In the example (Item 2) the correct tool is the one presented in the middle (solution: put the loop over the straw, bend the straw over and lift). **(B) Familiar Tools Test:** In the familiar tools example (Practice Item 0.2) the spatula is the correct tool to take the fried egg out of the pan (shove the spatula underneath the egg and lift it out of the pan).

patient, e.g., rather grip-formation or the movement-content or both. Patients were allowed to correct their movement in a second attempt if their first try was erroneous. If out of two attempts the second was improved, then the second attempt was evaluated for the 4-point Production Score. The maximum score that could be achieved was 10 points for the Selection Scale, 10 points for the Execution Scale (each 5 items \times 2 points) and 20 points for the 4-point Production Score (5 items \times 4 points).

2.3. Validation in a Naturalistic Action Task

In addition, it was tested how patients performed in a tool-use activity that involves several steps. Inspired by the Naturalistic Action Test (NAT) by Schwartz, Segal, Veramonti, Ferraro, and Buxbaum (2002) the patients' task was to prepare breakfast. In the original version patients were instructed to prepare a cup of coffee with milk and sugar and a slice of toast with butter and jam. For practical and cultural reasons (to avoid spilling accidents with fluid milk and because Germans do not typically use the alternative milk powder) in the current NAT participants were asked to make a cup of tea with sugar and a toasted slice of bread with butter and jam. Participants had all the time they needed and all ingredients and devices were placed in front of them in a defined setting (see Supplementary Fig. S.1). No distractor tools were presented. This task is more complex than the FTT, because in addition to applying the tools and objects correctly, the patient has to memorize the instructions and arrange the task into accumulative steps using executive functions (Goldenberg, 2008, 2011; Schwartz, Buxbaum, Veramonti, Ferraro, & Segal, 2001). To evaluate the task two scores were generated: The Accomplishment Score reported whether all steps were successfully performed and the Error Score indicated how many errors were made during the execution (e.g., spreading butter and jam with finger or spoon, or omitting to stir the tea). The final NAT-Score was formed by a combination of the Accomplishment Score and the Error Score.

2.4. Classic assessment of limb apraxia

Two classic tasks were used to further assess limb apraxia: imitation of hand-gestures (meaningful and meaningless) and

pantomiming the use of familiar objects. For imitation, the patients had to imitate ten meaningful and ten meaningless gestures. Each imitation test was started with one practice trial. The meaningless gestures were obtained from Goldenberg (1996). The set of meaningful gestures is a novel assembly and tests the imitation of emblems, which are gestures with conventional defined forms and meanings like “salute”, “listen carefully” or “blow somebody a kiss”. The division into the two categories was confirmed by asking 25 healthy volunteers to rate the familiarity of the gestures. Every meaningless gesture was rated by at least 72.0% of the volunteers as meaningless, for meaningful gestures every gesture was correctly identified with its meaning by at least 80.0% of the volunteers.

Both subtests were evaluated in the same way: per item participants received 2 points if their first attempt demonstrated the correct gesture and one point when the second attempt was correct, otherwise the experimenter indicated a *total error* (0 points). The gesture was evaluated as soon as the participant held it in a particular position. If the first held position was erroneous, the participant was asked to improve the imitated gesture. This allowed the experimenter to rate the second attempt. A maximum of 20 points for each imitation category could be achieved (per imitation category: 10 items \times 2 points).

In the pantomime task the participant was presented with a picture of an object and was verbally asked to show the typical movement for the particular object as if the object would be in their hand (e. g. “Show me how to hit a nail with a hammer.”). Presenting a picture together with the verbal command facilitates task comprehension for aphasic patients. The test comprised up to three practice trials and 8 test items, which were taken from the pantomime task by Goldenberg, Hartmann, and Schlott (2003). Of the original task seven items were excluded because they appeared outdated and six items were excluded because of being the least sensitive. The evaluation of the pantomime execution was equivalent to the other tool-use tasks with an Execution Scale and a 3-point Production Score. In the Execution Scale it was evaluated, whether the first or second attempt was correct (*first correct* = 2 points, *second correct* = 1 point). 0 points were noted if the correct pantomime was not shown at the second attempt (*total error*). The maximum score was 16 points for the

Execution Scale (8 items \times 2 points). For the 3-point Production Score, the performance per item was rated for the three parameters grip-formation, movement-content and movement-orientation. Accordingly, if an item was solved correctly three points were achieved leading to a maximum score of 24 points for the 3-point Production Score (8 items \times 3 points).

Additionally, the external apraxia-tests AST (Vanbellingen, 2012, 2013) and KAS (Weiss et al., 2013) were tested to gain information about concurrent validity for our imitation and pantomime scores.

2.5. Neuropsychological assessment

In both patient groups, motor and language skills as well as semantic processing were further examined to determine divergent validity. A short version of the Wolf Motor Function Test (WMFT; Wolf et al., 2001), the “Token” subtest of the Aachener Aphasia Test (AAT; Huber, Poeck, Weniger, & Willmes, 1983) and subtests of the Bogenhausen Semantic Assessment (BOSU; Glindemann, Klintwort, Ziegler, & Goldenberg, 2002) were applied.

The WMFT is used to test the contralesional arm and hand function with eight items. The AAT Subtest “Token Test” evaluates language comprehension. The BOSU was applied to test semantic knowledge with four object related picture based subtests (assigning objects to a situation, sorting objects according to a. primary or b. secondary semantic characteristics and sorting objects by color).

The Star Cancellation and Line Bisection Test (Plummer, Morris, & Dunai, 2003) assessed neglect. Care was taken to avoid an influence of visuo-spatial neglect on the tool-use assessment. If a visuo-spatial neglect or hemianopia was diagnosed, the item-set of the Familiar and NTT was shifted towards the unaffected hemi-space. Further, before the start of a subtest all participants were instructed once to pay attention to each of three tools, but for patients with visuo-spatial deficits this request was repeated for each trial.

2.6. Test applicability and acceptance

Test applicability and acceptance were estimated by questioning six occupational therapists and a subgroup of patients (LBD: $N = 26$, RBD: $N = 17$) with the test-acceptance questionnaire AKZEPT-L adapted from Kersting (2008).

2.7. Statistical analyses

All behavioral analyses were conducted in IBM SPSS Statistics 24.

Normative data. To obtain Cut-Off values, the 5th percentile was computed for each test based on the performance of the normative HC group ($N = 82$).

Psychometric data. For the analysis of patient performance, the total scores per task and scale were used. To compare scales with different maximum scores, percentage scores for all scales were computed.

Because no variables were normally distributed (tested with χ^2 , $p = .000$), non-parametric tests were used. For correlations (i.e., inter-test correlations, interrater reliability, concurrent and divergent validity measures) Kendall's Tau

was applied. For calculating differences between scales (Selection, Execution and the 4-point Production Score) the Wilcoxon Test was used. Internal consistency was computed using a conventional version of congeneric reliability (Composite Reliability) (Cho, 2016, p. 664). Criterion validity was determined by group comparisons using Mann–Whitney-U tests. Differences between groups should demonstrate typical laterality results found in the literature (for a similar approach see e.g., Vanbellingen et al., 2010). For concurrent validity we computed correlations with the AST (Vanbellingen, 2012, 2013) and KAS tests (Weiss et al., 2013). In order to obtain a comparable measure for the different tests, first a combined Imitation and Pantomime Score was computed. This accommodates the less specific evaluation methods used by the AST and KAS, respectively. For calculations including the AST, our score was computed as follows: (Imitation Meaningful + Imitation Meaningless + Pantomime Execution)/3. For calculations including the KAS our data was summarized as follows: (Imitation Meaningful + Imitation Meaningless + Pantomime Production)/3. To allow for computations of the less specific evaluation of the AST with KAS values, KAS data was transformed into dichotomous answers for every item.

3. Results

3.1. Normative data

Cut-Off values were achieved by testing 82 HC persons (HC group). For Cut-Off values the 5th percentile of the performances by the HC group was determined. All Cut-Off values of real tool-use tasks are listed in Table 2. In contrast to the NTT, Cut-Off values in the FTT, both in selection and application were found to be close to ceiling. In our NAT Breakfast Task adapted from Schwartz et al. (2002) we exchanged coffee making by tea cooking. Nevertheless, the same Cut-Off value was found.

In addition, normative data for imitation and pantomime tasks were collected. For imitation of meaningless and meaningful gestures each a maximum of 20 points could be achieved. Different Cut-Off values need to be considered for age groups 21–50 years (meaningless: 16, meaningful: 18 points) versus 51–80 years (meaningless: 15, meaningful: 16 points). For the pantomime task the Cut-Off values are 12 points in the Execution Scale (Max. 16 points) and 22 points in the 3-point Production Score (Max. 24 points).

3.2. Psychometric data

Task related comparisons and correlations were run for the LBD group only ($N = 33$) to ensure sufficient variability in performance.

3.2.1. Interrater reliability

The interrater-reliability was determined based on data of 15 HC persons (18.3%) rated by I. B. and C. H.-W. (novice) as well as data of ten LBD patients (27.8%) evaluated by I.B., P.L. and I.M. (the latter two were novices). Novice raters used a manual for evaluation guidance. For HC, a correlation was not

Table 2 – Cut-Off values for tool-use apraxia relevant tasks.

Test	Scale	Age/sex	Max. score	Cut-Off value
Familiar Tools Test	Selection Scale	21–50	10	9
		51–80		8
	Execution Scale	10	9	
Novel Tools Test	4-Point Production Score		20	20
	Selection Scale		10	6
		Execution Scale		Female
		Male	10	7
	4-Point Production Score	Female	20	17
Male		18		
NAT Breakfast Task	NAT-Score		6	4

analyzed because of missing variance in data, instead only percent of agreement is reported. The mean agreement in the subtests exceeded 95% (FTT Selection: 97.3%, FTT Execution: 98.7%, FTT 4-point Production: 100%, NTT Selection: 94.7%, NTT Execution: 85.0%, NTT 4-point Production: 96.0%).

Substantial to high interrater reliability was found for the evaluation of LBD patients' performance in the FTT and NTT subtests (Kendall's Tau correlation coefficient FTT Selection $\tau = .965$, $p = .002$, FTT Execution $\tau = .912$, $p = .002$, FTT 4-point Production $\tau = 1.000$; NTT Selection $\tau = .577$, $p = .038$, NTT Execution $\tau = .786$, $p = .004$, NTT 4-point Production $\tau = .735$, $p = .012$). There was 100% interrater congruence for diagnosing a patient as apraxic versus non-apraxic for each of the subscales.

3.2.2. Internal consistency and intercorrelations

Internal consistency was high in all subscales of the FTT ($CR \geq .768$) and the NTT ($CR \geq .742$). Please note, for one single NTT item (Item No 1) patients in our sample had no difficulties solving the correct action latest at the second attempt. While at the same time this item proved to be highly sensitive for selection errors, there was no variability in the NTT 4-point Production Score. Missing variability in one item prevented us from running internal consistency computations for the 4-Point Production Score. To still provide an estimate for the remaining items of this scale ($CR = .848$), NTT item No 1 was removed only for the NTT 4-point Production Score.

Intercorrelations were measured by correlating subtests of the here tested limb apraxia battery in the LBD group. Correlations between subscales of FTT and NTT were significant for almost all subscales ($\tau \geq .338$, $p \leq .014$). The same tendencies of worse performance in one scale going along with impaired behavior in another scale were found for FTT Selection and NTT Selection ($\tau = .240$, $p = .092$) and FTT 4-point Production Score and NTT Selection Scale ($\tau = .260$, $p = .070$). Imitation and Pantomime tasks and subscales correlated significantly with each other ($\tau \geq .393$, $p \leq .003$). While pantomime of tool-use and real tool-use tasks correlated consistently across scales ($\tau \geq .309$, $p \leq .025$) more variable associations were found between imitation tasks and the different subscales of the real tool-use tasks (Min: $\tau = .070$, $p = .601$, Max: $\tau = .441$, $p = .001$). For a detailed correlation table please see [Table S.1](#) (supplementary material).

[Fig. 2](#) demonstrates that the different domains correlate, but that patients score worse on the classic apraxia tests versus the real tool-use tasks. In the LBD group, imitation and

pantomime tasks were more sensitive than the real tool-use tasks (LBD group: $Z = -3.026$, $p = .002$). In the RBD group the trend was in the same direction ($Z = -1.682$, $p = .105$). For the classic imitation and pantomime tests 35–79% of the patients were impaired according to the normative data (LBD: 78.8%; $n = 22$; RBD: 35.0%; $n = 7$), while in real tool-use fewer but still an impressive number of patients showed difficulties in at least one subscale (LBD: 57.6%; $n = 19$; RBD: 30.0%; $n = 6$).

3.2.3. Concurrent validity

Concurrent validity could not be determined appropriately due to a lack of comparable other tests with actual tools and objects providing psychometric information and/or normative data. However, we computed correlations for the here used imitation and pantomime scores with the external tests AST ($N = 33$) and KAS ($N = 14$). The AST and KAS only provided a combined Cut-Off value to diagnose apraxia for the classic tests imitation of gestures and pantomime of tool-use. Accordingly, from the current test battery imitation and pantomime subtests were combined as well. Inter-test correlations of the combined measures were highly positive (DILA-S & AST: $\tau = .500$, $p = .000$; DILA-S & KAS: $\tau = .522$, $p = .010$; AST & KAS: $\tau = .530$, $p = .012$).

Further, we analyzed the association between the performance on the FTT and NTT subscales with the performance on a Naturalistic Action Task (NAT: making breakfast) adapted for this study. While FTT subscales correlated significantly with the NAT Breakfast Task ($\tau \geq .414$, $p \leq .006$), NTT subscales did not correlate with the NAT Breakfast Task ($\tau \leq .190$, $p \geq .179$).

3.2.4. Divergent validity

In order to evaluate the influence of external variables on test performance divergent validity was analyzed for motor, visuospatial and language tests in LBD patients. Motor ($\tau \leq .272$, $p \geq .058$) and visuospatial measures ($\tau \leq .218$, $p \geq .162$) did not correlate significantly with any of the tool-use subscales besides the NTT Selection Scale (WMFT: $\tau = .285$, $p = .036$; Neglect tests: $\tau = -.321$, $p = .035$). To further elucidate these effects, group comparisons were run. Accordingly, tool-use performance in patients with difficulties in the tested motor ($N = 21$) or visuospatial tasks ($N = 4$) did not differ from patients without motor ($N = 12$) or visuospatial deficits ($N = 29$) in all tool-use scales ($U \leq 116.0$, $p \geq .109$) except for the NTT Selection (WMFT: $U = 53.0$, $p = .005$; Neglect tests: $U = 18.5$, $p = .025$).

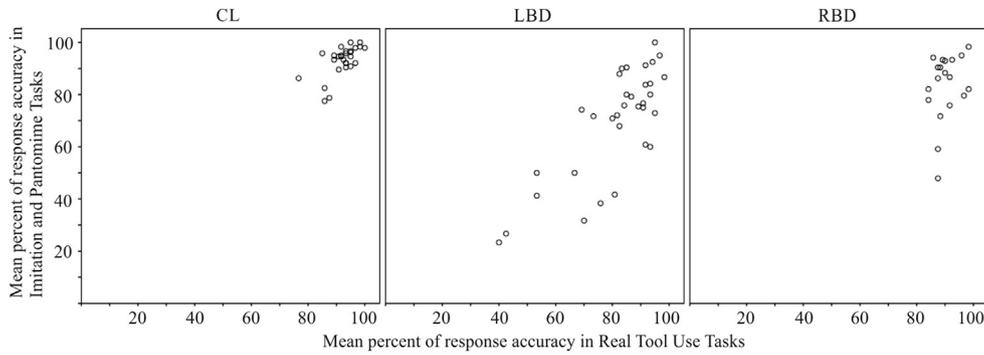


Fig. 2 – Scatter plot for mean percent of accuracy in imitation and pantomime tasks (Y-axes) and Novel and Familiar Tools Tests (X-Axes) for CL (N = 28), LBD (N = 33) and RBD (N = 20) groups.

Language comprehension correlated highly with all tool-use subscales ($\tau \geq .317, p \leq .036$). To further elucidate the influence of language on tool-use deficits, group comparisons were run. Compared to patients without deficits ($N = 10$), patients with language comprehension deficits ($N = 23$) demonstrated worse performance in all tool-use scales ($U \geq 23.5, p \leq .048$) besides NTT 4-point Production ($U = 75.0, p = .108$).

Further, the semantic processing task (BOSU) correlated in almost all subtests with the FTT but not for most of the NTT subtests (FTT: $\tau \leq -.364, p \leq .031$, NTT: $\tau \geq -.223, p \geq .159$). The exceptions are as follows: the BOSU subtest sorting objects by color did not correlate with FTT ($\tau \geq -.187, p \geq .191$) and sorting objects by secondary semantic characteristics correlated with NTT ($\tau = -.329, p = .038$).

3.2.5. Criterion validity

According to the majority of studies in the literature limb apraxia in right handed stroke patients is a disorder that typically occurs in 30–50% of the patients with left hemisphere damage and 0 to 30% in patients with right hemisphere damage (Donkervoort et al., 2000). Therefore, we consider a strong left lateralization in tool-use performance as evidence for criterion validity, i.e., if LBD patients score worse on all

apraxia tests compared to healthy persons (for a similar approach see e.g., Roy, Square-Storer, Hogg, & Adams, 1991; Salazar-Lopez, Schwaiger, & Hermsdoerfer, 2016; Vanbellingen et al., 2010). Since RBD patients typically are not as strongly affected, significant differences should not be expected between RBD patients and the HC group (e.g., Roy et al., 1991; Salazar-Lopez et al., 2016; Vanbellingen et al., 2010).

Group comparisons are displayed in Fig. 3. The LBD group scored significantly worse in most tool-use measures compared to the CL group ($U \leq 306.0, p \leq .021$), except for the selection of familiar tools ($U = 365.5, p = .104$) and the qualitative assessment of the application of novel tools (4-point Production: $U = 359.0, p = .110$). Despite some LBD patients being identified as apraxic in the latter two scales, the group comparison only demonstrated a tendency towards worse performance of the LBD group.

In total, out of the 33 LBD patients 57.6% ($n = 19$) were classified as apraxic in real tool-use according to our normative data. Behavior was defined apraxic if a patient scored below Cut-Off in at least one of the subscales of NTT or FTT.

Except for the application of familiar tools the LBD and RBD group did not differ ($U \geq 259.0, p \geq .188$). For the Execution Scale and 4-point Production Score in the FTT the

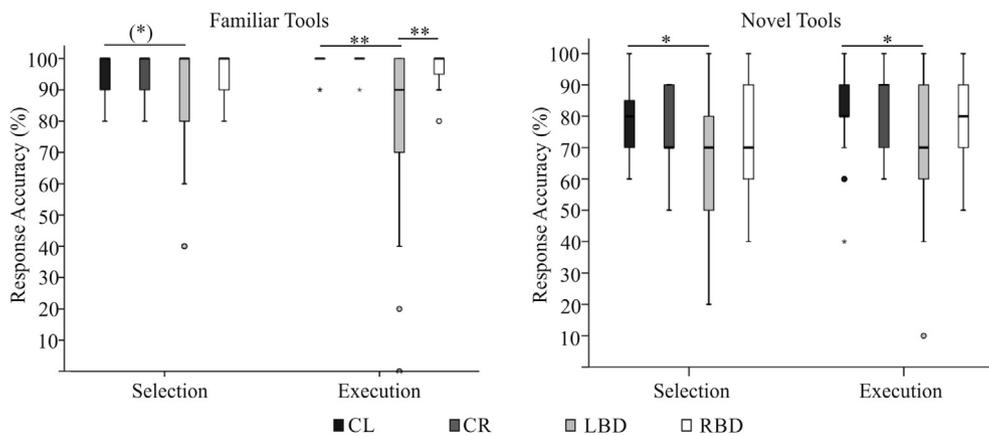


Fig. 3 – Boxplots present the accuracy (%) for the Selection and Execution Scales in the Familiar Tools and Novel Tools Test for each group (CL: N = 28, black; CR: N = 18, dark grey; LBD: N = 33, light grey; RBD: N = 20, white). Significant differences between groups are indicated [$p \leq .010, *p \leq .050, (*) p < .100$].**

LBD group performed significantly worse than the RBD Group (Execution: $U = 179.0$, $p = .005$; 4-point Production: $U = 222.0$, $p = .009$).

Despite the lack of group differences between RBD patients and their respective HC group ($U \geq 130.0$, $p \geq .089$), some individuals demonstrated difficulties in the NTT and/or in the FTT. Based on our normative data, 6 RBD patients (30.0%) were classified as apraxic in at least one tool-use measure (for details see [supplementary Table S.3](#)).

3.2.6. Apraxic profile

In order to deliver further insight into the measured construct, this section includes a brief description of the dysfunction in the applied tool-use tests that was demonstrated by the tested sample of unilateral stroke patients. Detailed information can be found in the supplementary material.

We found performance differences between the tool-selection and tool-application parameters.

Performance varied across the different scales on a group level (see also [Fig. S.2](#)). For example, in patients with LBD, the NTT Selection and Execution were similarly affected, while in the FTT the Execution was performed worse compared to Selection. The latter difference was not found for RBD patients. We found a few individuals to demonstrate selective deficits in NTT or FTT, respectively. In both tool-use tests deficits in Selection and Execution emerged most often combined, but selective deficits were demonstrated in a few individuals. For a detailed description on selectivity, please see [Tables S.2 and S.3](#) and the corresponding text in the supplementary material.

Descriptive information on the frequency of error-types occurring in the applied tests is available (see also [Fig. S.3](#)). When looking at the parameter (grip-formation, grip-orientation, movement-content or movement-orientation) for which errors were produced most commonly, in both LBD and RBD groups, the movement-content was most frequently affected. In patients with LBD, in the FTT, errors appeared for all evaluated parameters. However, grip-formation and orientation of thumb were relatively error-resistant, particularly in the NTT.

Further, the supplementary material provides guiding values that indicate the presence of mild, moderate and severe tool-use apraxia in order to enable a severity differentiation. The descriptive data demonstrated that the use of familiar tools caused most severe difficulties, especially in patients with LBD.

3.2.7. Test applicability and acceptance

We collected feedback on the DILA-S by handing a questionnaire to local occupational therapists (adapted from [Kersting, 2008](#)), who applied the developed tests. The DILA-S received the overall grade “good” (Mean = 1.83 on German school grades scale). The therapists estimated face validity (Mean = 1.50), reliability (Mean = 1.50) and applicability with respect to the workload for implementing the test battery (Mean = 2.08) to be “good”. Additionally, patients (LBD: $N = 26$, RBD: $N = 17$) stated that all tasks were easy to understand and that they were not overstrained by the testing situation.

4. Discussion

We introduced two new tasks – the FTT and the NTT – to diagnose limb apraxia for tool-use including actual tools and objects. In addition, one naturalistic multistep action was evaluated. Next to providing normative data based on a healthy sample ($N = 82$), psychometric data of patients with first unilateral left or right hemisphere stroke were reported. In the following we first will evaluate the reliability and validity outcomes including a discussion of the applicability of the DILA-S. Thereafter we will discuss the role of actual tool-use as diagnostic instrument.

4.1. Psychometric data

4.1.1. Interrater-reliability and consistency measures

To evaluate the tests' reliability, we computed interrater-reliability and internal consistency measures. Further, we computed inter-correlations for the tests and their subscales.

Substantial to high interrater-reliability was achieved. Importantly, the different rater agreed completely for diagnosing a patient as apraxic versus non-apraxic for each of the subscales. Correlations of performance scores were significant for all subscales. For the Selection component, a satisfactory agreement had to be expected, since for each item the correct tool is unequivocally determined. Whereas the qualitative judgment of the tool-application (Execution Scale) provides room for variability between raters. However, for the here described test evaluation, sheets were used that provide clearly defined action characteristics that had to be fulfilled in order for the action to be classified as correct. The description of each action follows the parameters of the 4-point Production Score: grip-formation, grip-orientation, movement-content and movement-orientation. To maintain high agreement between different judges for the evaluation of the Execution Scale, it is therefore recommended to make use of this qualitative guide delivered with the supplementary material.

We found internal consistency with values larger than .7 to be high. Further, nonparametric correlations computed for the purpose of estimating construct consistency within the subdomains of the battery (i.e., meaningful and meaningless imitation, pantomime, novel and familiar tool-use) revealed predominantly significant associations. Correlations were found within tool-use domains and within classic tests. Inter-correlations between most of the subscales of the FTT and NTT were significant, except, the Selection Scale in the FTT and NTT did only show a weak association. This suggests that the selection of applicable novel and familiar tools may only partly share an underlying construct. Further, performance on the classic imitation and pantomime tests only partly correlated with tool-use. While pantomime of tool-use and real tool-use tasks correlated consistently, more variable associations were found between imitation and real tool-use tasks. In addition, in line with previous literature our data clearly showed worse performance for LBD and RBD patients in pantomime and imitation tasks in comparison to real tool-use tasks ([De Renzi et al., 1982](#); [Randerath et al., 2011](#)). It needs to be noted, that we did not expect perfect correlations between

the subtests because we propose that pantomime, imitation and real tool-use tests may share underlying motor cognitive abilities, but are considered to test essentially different domains of this construct. Instead, as expected these imperfect associations between the subdomains suggest that assessing only single domains may risk obtaining an incomplete picture.

Future studies including larger samples than the current may provide a better insight into the association or distinction between subdomains of apraxia measured with the here introduced tests.

4.1.2. Indicators of validity

In order to further elucidate the measured construct, we provided a rough estimation of concurrent validity by computing correlations with external measures. Until now there is a lack of equivalent external tests assessing the selection and application of actual tools and objects. Here, an assessment of concurrent validity was approximated by computing correlations between tests of a closely related construct: Patient performance in the here used imitation and pantomime scores of the DILA-S correlated significantly with corresponding results collected with the external tests AST (Vanbellingen, 2012, 2013) and KAS (Weiss et al., 2013). This estimation needs to be interpreted cautiously, especially given the above described differential findings between the classic tests and real tool-use.

In addition, we found that performance in FTT but not NTT correlated significantly with performance on the Naturalistic Action Task (NAT: making breakfast) adapted for this study. Making breakfast is a sequential task that requests general abilities such as planning, working memory and attention (Buxbaum, Schwartz, & Montgomery, 1998; Schwartz et al., 1998). But the task is also highly familiar and includes the selection and application of known objects and tools from a rather cluttered environment in which intact mechanical problem solving may not suffice to solve the task. The strong association with the FTT but not the NTT is therefore not surprising.

While this provides some positive indication of concurrent validity, in the future this property should be further assessed using matching test approaches.

Divergent validity was evaluated by analyzing the influence of external variables on test performance. Whereas the assessed motor and visuospatial measures seemed to have little impact on performance in our tool-use tests, a difference was found based on whether language comprehension was affected. A greater impairment in speech comprehension went along with more severe tool-use deficits. A self-evident explanation may be the partly overlapping neural substrates of language and praxis. Previous lesion studies reported that worse performance in the Token Test as well as worse performance in pantomime of tool use and imitation of meaningless gestures go along with overlapping lesions in the parietal lobe (Goldenberg & Randerath, 2015).

The underlying test construct may further emerge by delimiting performance on FTT and/or NTT from performance on tasks that are thought to be related in a dissociable way. For example, whereas correctly solving the FTT is assumed to at least partly rely on semantic knowledge (Goldenberg & Spatt, 2009; Hodges et al., 1999; Randerath et al., 2011) performing

the NTT is thought to be predominantly based on mechanical problem solving, which presupposes familiarity with general principles of physics and mechanics rather than reliance on knowledge of familiar objects (Hegarty, 2004; Povinelli, 2000; Zago & Lacquaniti, 2005). Accordingly, in patients with LBD, performance in the semantic processing task (BOSU) showed significant correlations with the achieved scores in all FTT subscales and no or only weak correlations with the NTT subscales. However, semantic knowledge appears not to be the sole medium ensuring appropriate familiar tool-use. For example, three LBD patients with difficulties in the FTT were not impaired in the semantics test but did show difficulties with the NTT Selection. This supports previous findings by for example Hodges et al. (1999) as well as by Goldenberg and Spatt (2009) who suggested that the selection and application of familiar tools likely is supported by both retrieval of functional knowledge from semantic memory and mechanical problem solving. A possible factor explaining the high coincidence of these co-occurring deficits going along with the FTT may be lesion size. A similar explanation may hold for the finding that impaired performance in the NTT Selection as the only scale went along with motor and visuo-spatial deficits.

Future studies with the DILA-S including larger patient groups may consider applying a factor analysis to further elucidate the existing interrelations or differentiating factors. In this context, it should be considered to include an additional task unrelated to tool-use or semantics that measures a construct closely related to mechanical problem solving.

Left lateralization going along with limb apraxia in right handed patients is indeed a long established finding (Liepmann, 1905). Previous literature reports a prevalence of limb apraxia in 30–50% of the patients with left hemisphere damage and 0 to 30% in patients with right hemisphere damage (Donkervoort et al., 2000). Accordingly, results of left lateralization in our tests were interpreted as an indicator for criterion validity (for a similar approach see e.g., Roy et al., 1991; Salazar-Lopez et al., 2016; Vanbellingen et al., 2010). Our data is in line with these approximations, although indicating prevalence at the higher end. Based on our normative data, 58% of LBD and 30% of RBD patients were classified as apraxic in at least one tool-use measure. It needs to be mentioned that all RBD patients demonstrated difficulties in only one of the tool related subscales. Our group comparisons confirmed the expected laterality effects. While the patient group with LBD scored worse on almost all real tool-use scales compared to their respective HC group, the performance of the patient group with RBD did not differ from their respective HC group in all real tool-use related subscales. This suggests that lesions in the right hemisphere may more generally lead to preponderant mild forms of tool-use apraxia (in patients with right hand dominance, but see also Goldenberg (2013b)). However, on a group-level, RBD and LBD patients did not differ from each other, apart from the application of familiar tools.

Summarized, although apraxia of tool-use after RBD seems not to be rare, it still is less prevalent and less severe compared to the deficits after LBD. At the same time our results highlight that although limb apraxia may occur not as frequent or severe after RBD it is tenuous to omit RBD patients in limb apraxia diagnostics in clinical settings as well as limb apraxia studies

(see also Bienkiewicz, Brandi, Hughes, Voithl, & Hermsdoerfer, 2015; Goldenberg, Hartmann-Schmid, Surer, Daumüller, & Hermsdoerfer, 2007; Hartmann, Goldenberg, Daumüller, & Hermsdoerfer, 2005; Rumiati, 2005; Salazar-Lopez et al., 2016; Schwartz et al., 1999; Schwartz et al., 1998; Wirth et al., 2016).

Acceptance of the DILA-S seemed to be appropriate amongst applicants. Patients stated that all tasks were easy to understand and that they were not overstrained by the testing situation. Local therapists estimated face validity, reliability to be “good”. Its applicability with respect to the workload for implementing the test instrument was judged to be “good” as well.

4.1.3. Apraxic profile

According to the normative data, the use of familiar tools appeared to cause most severe difficulties, especially in patients with LBD. When looking at the type of error that was produced most commonly across the evaluated parameters, both patient groups concur. In line with prior studies evaluating tool-use with single tools (Randerath, 2009; Randerath, Li, Goldenberg, & Hermsdoerfer, 2009), the movement-content parameter was most frequently affected. Grip-formation and orientation of thumb were relatively error-resistant in both tool-use tests, but particularly in the NTT. One needs to acknowledge that for all NTT items the handles look identical (including the practice trials), which may have promoted correct grasping.

Comparisons of performance between application and selection measures demonstrated that there can be dissociable performance in individuals and underlines that the use of the different tool-use tests and scales is informative. We suggest that within the clinical context it is worth to both consider a trial-and-error Execution Scale as well as a qualitative Production Score. The Execution Score may reveal whether there are actions that a patient fails at completely and may therefore render independence in daily life potentially impossible. The 4-point Production Score indicates whether and which single components are not performed correctly and thus may deliver implications for the approach of rehabilitation.

4.2. The role of real familiar and novel tool-use

For many potential applicants, the ecological validity of limb apraxia assessments may not be obvious. Many studies unraveling limb apraxia have applied tasks of the gesture domains (Barde, Buxbaum, & Moll, 2007; Goldenberg & Randerath, 2015; Roy et al., 1991; Vanbellingen et al., 2011; Weiss et al., 2013). But only few addressed the domain of real tool-use (e.g., Goldenberg & Spatt, 2009; Leiguarda et al., 2014), despite the fact that the latter apraxia domain obviously is an important player in managing activities of daily living. Here we provided an approach to test real tool-use that is applicable in clinical settings, as well as of interest for scientific purposes.

The phenomenon of patients being better at real tool-use tasks in clinical or research settings may be another reason for why real tool-use tasks are neglected for standard diagnostic. Real tool-use tasks in an experimental setting typically are far more organized compared to home settings and therefore may facilitate the use of tools and objects. Accordingly, we replicated the finding that imitation and pantomime

tasks are more sensitive to erroneous behavior. However, we also demonstrated that real tool-use can be affected in a large number of patients with left and RBD. It thus should not be underestimated.

Further, we showed that impaired tool selection and application often co-occur, and in some individuals even can be affected selectively. Both, appropriate tool selection and application, are important to manage daily life in home setting environments. We therefore advise against dropping either evaluation since important information may be lost.

At last, limb apraxia rehabilitation developments are sparse and stagnate since several years (Buxbaum et al., 2008; Cantagallo, Maini, & Rumiati, 2012). Together with other studies claiming that the use of tools and objects is especially relevant to empower patients for an independent daily live (De Renzi & Lucchelli, 1988; Goldenberg & Hagmann, 1998a; Mayer et al., 1990; Poeck, 1983) the current data underlines the importance of taking tool-use apraxia seriously.

4.3. Future directions

Overall, our results tune optimistic that the DILA-S is a useful instrument for standard use in clinical settings. The questioned occupational therapists suggested that standard use in the daily clinical routine will be facilitated by providing a very detailed test manual as well as a hands-on training. The manual will soon be available online. Efforts for translation in English started. However, despite all the profound reasons, it still may take time and effort until limb apraxia diagnostics will be accepted as standard instruments in clinics. Several other factors beyond the scope of this work may play a role as well, such as straightening out within each clinic whether its diagnosis should be allocated to occupational therapy or neuropsychology.

Next to diagnosing limb apraxia, the DILA-S may be a useful instrument to measure improvement of tool-use deficits during the course of rehabilitation. For this purpose, ideally parallel tests using different items will be developed.

Evidence compiles that there exist different routes in the brain from the visual system to fronto-parietal regions that guide and support the planning and production of actions. The finding of single cases with selective deficits in the FTT and NTT fits well with the theory of differential processing routes (Milner & Goodale, 2008) which has been discussed in several studies with respect to the praxis system (Hodges et al., 1999; Osiurak et al., 2009; for a review see; Reynaud, Lesourd, Navarro, & Osiurak, 2016) and substantiated by several imaging studies (Barde et al., 2007; Binkofski & Buxbaum, 2013). The subtypes or domains of limb apraxia may evolve due to the deficient retrieval and integration of differing sources of information distributed in this dual route system (Randerath, 2009). Within the framework of these route models it needs further clarification how the selection and execution components are distributed for familiar versus novel tool-use.

Conflict of interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no

significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We further confirm that any aspect of the work covered in this manuscript that has human patients has been conducted with the ethical approval of all relevant bodies and that such approvals are acknowledged within the manuscript.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). She is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from J_Randerath@hotmail.com.

Acknowledgement & Funding

This work was supported by grants of the DFG (grant no. RA2492/3-1) and the EU FP7 Marie Curie Zukunftscolleg Incoming Fellowship Programme at the University of Konstanz (grant no. 291784) and an internal Zukunftscolleg Co-Funding grant to support research transfer awarded to Dr. Jennifer Randerath as well as the *Landesgraduiertenförderungsgesetz* (LGFG) in Baden-Württemberg awarded to M. Sc. Ilka Buchmann.

We thank Celina Hirth-Walter, Patrick Lipfert and Inara Makhkamova for their interrater judgments, Prof. Dr. Joachim Liepert, Imke Büsching, Dr. Thomas Hassa, Stefan Spiteri and Lisa Friedrich-Schmieder for their support with administrative questions in the rehabilitation center of the Kliniken Schmieder, as well as all participants.

All partners had no influence on the conduct of research, preparation of the article, study design, analysis and interpretation of data, writing the report and/or decision to submit the article for publication.

Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.cortex.2017.06.001>.

REFERENCES

- Barde, L. H., Buxbaum, L. J., & Moll, A. D. (2007). Abnormal reliance on object structure in apraxics' learning of novel object-related actions. *Journal of the International Neuropsychological Society*, 13, 997–1008.
- Belanger, S. A., Duffy, R. J., & Coelho, C. A. (1996). The assessment of limb apraxia: An investigation of task effects and their cause. *Brain and Cognition*, 32, 384–404.
- Bienkiewicz, M. M., Brandi, M. L., Hughes, C., Voitl, A., & Hermsdoerfer, J. (2015). The complexity of the relationship between neuropsychological deficits and impairment in everyday tasks after stroke. *Brain and Behavior*, 5.
- Binkofski, F., & Buxbaum, L. J. (2013). Two action systems in the human brain. *Brain and Language*, 127, 222–229.
- Burrell, J. R., Hornberger, M., Vucic, S., Kiernan, M. C., & Hodges, J. R. (2014). Apraxia and motor dysfunction in corticobasal syndrome. *Plos One*, 9, e92944.
- Buxbaum, L. J., Haaland, K. Y., Hallett, M., Wheaton, L., Heilman, K. M., Rodriguez, A., et al. (2008). Treatment of limb apraxia: Moving forward to improved action. *American Journal of Physical Medicine & Rehabilitation*, 87, 149–161.
- Buxbaum, L. J., Schwartz, M. F., & Montgomery, M. W. (1998). Ideational apraxia and naturalistic action. *Cognitive Neuropsychology*, 15, 617–643.
- Cantagallo, A., Maini, M., & Rumiati, R. I. (2012). The cognitive rehabilitation of limb apraxia in patients with stroke. *Neuropsychological Rehabilitation*, 22, 473–488.
- Chandra, S. R., Isaac, T. G., & Abbas, M. M. (2015). Apraxias in neurodegenerative dementias. *Indian Journal of Psychological Medicine*, 37, 42–47.
- Cho, E. (2016). Making reliability reliable: A systematic approach to reliability coefficients. *Organizational Research Methods*, 19, 651–682.
- Cotelli, M., Manenti, R., Brambilla, M., & Balconi, M. (2014). Limb apraxia and verb processing in Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 36, 843–853.
- De Renzi, E., Faglioni, P., & Sorgato, P. (1982). Modality-specific and supramodal mechanisms of apraxia. *Brain*, 105, 301–312.
- De Renzi, E., & Lucchelli, F. (1988). Ideational apraxia. *Brain*, 111(Pt 5), 1173–1185.
- Deutsch-Lezak, M., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). Construction and motor Performance: Examining for apraxia. In M. Deutsch-Lezak, D. B. Howieson, E. D. Bigler, & D. Tranel (Eds.), *Neuropsychological assessment* (5th ed., pp. 607–609). New York: Oxford University Press.
- Donkervoort, M., Dekker, J., Van Den Ende, E., & Stehmann-Saris, J. (2000). Prevalence of apraxia among patients with a first left hemisphere stroke in rehabilitation centres and nursing homes. *Clinical Rehabilitation*, 14, 130–136.
- Dovern, A., Fink, G. R., & Weiss, P. H. (2011). Diagnostik und Therapie der Gliedmaßenapraxie. *Fortschritte Neurologische Psychiatrie*, 79, 345–357.
- Glindemann, R., Klintwort, D., Ziegler, W., & Goldenberg, G. (2002). *Bogenhausener Semantik-Untersuchung BOSU* (Vol. 1). Jena: Urban & Fischer Verlag München.
- Goldenberg, G. (1996). Defective imitation of gestures in patients with damage in the left or right hemispheres. *Journal of Neurology*, 61, 176–180.
- Goldenberg, G. (2008). Apraxien. *Aktuelle Neurologie*, 35, 34–48.
- Goldenberg, G. (2011). *Apraxien*. Göttingen: Hogrefe Verlag GmbH & Co. KG.
- Goldenberg, G. (2013a). Apraxia. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4, 453–462.
- Goldenberg, G. (2013b). Apraxia in left-handers. *Brain*, 136, 2592–2601.
- Goldenberg, G. (2013c). *Apraxia: The cognitive side of motor control*. Oxford: Oxford University Press.
- Goldenberg, G., Daumüller, M., & Hagmann, S. (2001). Assessment and therapy of complex activities of daily living in apraxia. *Neuropsychological Rehabilitation*, 11, 147–169.

- Goldenberg, G., & Hagmann, S. (1998a). Therapy of activities of daily living in patients with apraxia. *Neuropsychological Rehabilitation*, 8, 123–141.
- Goldenberg, G., & Hagmann, S. (1998b). Tool use and mechanical problem solving in apraxia. *Neuropsychologia*, 36, 581–589.
- Goldenberg, G., Hartmann-Schmid, K., Surer, F., Daumüller, M., & Hermsdoerfer, J. (2007). The impact of dysexecutive syndrome on use of tools and technical devices. *Cortex*, 43, 424–435.
- Goldenberg, G., Hartmann, K., & Schlott, I. (2003). Defective pantomime of object use in left brain damage: apraxia or asymbolia? *Neuropsychologia*, 41, 1565–1573.
- Goldenberg, G., & Randerath, J. (2015 Aug). Shared neural substrates of apraxia and aphasia. *Neuropsychologia*, 75, 40–49. <http://dx.doi.org/10.1016/j.neuropsychologia.2015.05.017> (Epub 2015 May 22).
- Goldenberg, G., & Spatt, J. (2009). The neural basis of tool use. *Brain*, 132, 1645–1655.
- Hanna-Pladdy, B., Heilman, K. M., & Foundas, A. L. (2003). Ecological implications of ideomotor apraxia: Evidence from physical activities of daily living. *Neurology*, 60, 487–490.
- Hartmann, K., Goldenberg, G., Daumüller, M., & Hermsdoerfer, J. (2005). It takes the whole brain to make a cup of coffee: The neuropsychology of naturalistic actions involving technical devices. *Neuropsychologia*, 43, 625–637.
- Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, 8, 280–285.
- Hermsdoerfer, J., Li, Y., Randerath, J., Roby-Brami, A., & Goldenberg, G. (2013). Tool use kinematics across different modes of execution. Implications for action representation and apraxia. *Cortex*, 49, 184–199.
- Hodges, J. R., Spatt, J., & Patterson, K. (1999). “What” and “how”: Evidence for the dissociation of object knowledge and mechanical problem-solving skills in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 9444–9448.
- Huber, W., Poeck, K., Weniger, D., & Willmes, K. (1983). AAT–Aachener Aphasie test. Göttingen: Hogrefe.
- Kersting, M. (2008). Zur akzeptanz von Intelligenz- und Leistungstests. *Report Psychologie*, 33, 420–433.
- Leiguarda, R., Clarens, F., Amengual, A., Drucaroff, L., & Hallett, M. (2014). Short apraxia screening test. *Journal of Clinical and Experimental Neuropsychology*, 36, 867–874.
- Liepmann, H. (1905). Die linke Hemisphäre und das Handeln. *Münchener Medizinische Wochenschrift*, 48–49.
- Mayer, N., Reed, E., Schwartz, M. F., Montgomery, M., & Palmer, C. (1990). Buttering a hot cup of coffee: An approach to the study of errors of action in patients with brain damage. In D. Tupper, & K. Cicerone (Eds.), *The neuropsychology of everyday life: Assessment and basic competencies* (pp. 259–284). Boston Dordrecht London: Kluwer Academic Publishers.
- Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, 46, 774–785.
- Osiurak, F., Jarry, C., Allain, P., Aubin, G., Etcharry-Bouyx, F., Richard, I., et al. (2009). Unusual use of objects after unilateral brain damage: The technical reasoning model. *Cortex*, 45, 769–783.
- Osiurak, F., Jarry, C., Lesourd, M., Baumard, J., & Le Gall, D. (2013). Mechanical problem-solving strategies in left-brain damaged patients and apraxia of tool use. *Neuropsychologia*, 51, 1964–1972.
- Plummer, P., Morris, M. E., & Dunai, J. (2003). Assessment of unilateral neglect. *Physical Therapy*, 83, 732–740.
- Poeck, K. (1983). Ideational apraxia. *Journal of Neurology*, 230, 1–5.
- Poeck, K. (2006). Apraxie. In W. Hartje, & K. Poeck (Eds.), Vol. 6. *Klinische Neuropsychologie* (pp. 227–239). Stuttgart: Georg Thieme Verlag KG.
- Povinelli, D. (2000). *Folk physics for apes – the chimpanzee’s theory of how the world works*. Oxford, New York: Oxford University Press.
- Randerath, J. (2009). *Aspekte des Werkzeuggebrauchs: vom Greifen bis zur Ausführung*. Behaviorale und Neuronale Korrelate der Apraxie. Aachen: RWTH Aachen Universität.
- Randerath, J., Goldenberg, G., Spijkers, W., Li, Y., & Hermsdoerfer, J. (2011). From pantomime to actual use: How affordances can facilitate actual tool-use. *Neuropsychologia*, 49, 2410–2416.
- Randerath, J., Li, Y., Goldenberg, G., & Hermsdoerfer, J. (2009). Grasping tools: Effects of task and apraxia. *Neuropsychologia*, 47, 497–505.
- Reynaud, E., Lesourd, M., Navarro, J., & Osiurak, F. (2016). On the neurocognitive origins of human tool use: A critical review of neuroimaging data. *Neuroscience and Biobehavioral Reviews*, 64, 421–437.
- Roy, E. A., Square-Storer, P., Hogg, S., & Adams, S. (1991). Analysis of task demands in apraxia. *International Journal of Neuroscience*, 56, 177–186.
- Rumiati, R. I. (2005). Right, left or both? Brain hemispheres and apraxia of naturalistic actions. *Trends in Cognitive Sciences*, 9, 167–169.
- Salazar-Lopez, E., Schwaiger, B. J., & Hermsdoerfer, J. (2016). Lesion correlates of impairments in actual tool use following unilateral brain damage. *Neuropsychologia*, 84, 167–180.
- Salmaso, D., & Longoni, A. M. (1983). Hand preference in an Italian sample. *Perceptual and Motor Skills*, 57, 1029–1042.
- Schwartz, M. F., Buxbaum, L. J., Montgomery, M. W., Fitzpatrick-Desalme, E., Hart, T., Ferraro, M., et al. (1999). Naturalistic action production following right hemisphere stroke. *Neuropsychologia*, 37, 51–66.
- Schwartz, M. F., Buxbaum, L. J., Veramonti, T., Ferraro, M., & Segal, M. (2001). *NAT manual*. Retrieved from: <http://mri.org/wp-content/uploads/2016/01/NATManual.pdf>.
- Schwartz, M. F., Montgomery, M. W., Buxbaum, L. J., Lee, S. S., Carew, T. G., Coslett, H. B., et al. (1998). Naturalistic action impairment in closed head injury. *Neuropsychology*, 12, 13–28.
- Schwartz, M. F., Segal, M., Veramonti, T., Ferraro, M., & Buxbaum, L. J. (2002). The naturalistic action test: A standardised assessment for everyday action impairment. *Neuropsychological Rehabilitation*, 12, 311–339.
- Stamenova, V., Roy, E. A., Szilagy, G., Honjo, K., Black, S. E., & Masellis, M. (2015). Progression of limb apraxia in corticobasal syndrome: Neuropsychological and functional neuroimaging report of a case series. *Neurocase*, 21, 642–659.
- Tessari, A., Toraldo, A., Lunardelli, A., Zadini, A., & Rumiati, R. I. (2015). Stima: A short screening test for ideomotor apraxia, selective for action meaning and bodily district. *Neurological Sciences*, 36, 977–984.
- Unsal-Delialioglu, S., Kurt, M., Kaya, K., Culha, C., & Ozel, S. (2008). Effects of ideomotor apraxia on functional outcomes in patients with right hemiplegia. *International Journal of Rehabilitation Research*, 31, 177–180.
- Vanbellinggen, T. (2012). Apraxie: TULIA (test of upper limb apraxia) und AST (apraxia screen of TULIA). In K. Schädler, M. Lüthi, & W. Oesch (Eds.), *Assessments in der Rehabilitation – Band 1: Neurologie* (pp. 445–449). Bern: Verlag Hans Huber.
- Vanbellinggen, T. (2013). Assessment: Apraxia screen of TULIA – Gesten testen. *ergopraxis*, 6, 22–23.
- Vanbellinggen, T., Kersten, B., Van De Winckel, A., Bellion, M., Baronti, F., Muri, R., et al. (2011). A new bedside test of gestures in stroke: The apraxia screen of TULIA (AST). *Journal of Neurology, Neurosurgery, and Psychiatry*, 82, 389–392.
- Vanbellinggen, T., Kersten, B., Van Hemelrijk, B., Van De Winckel, A., Bertschi, M., Muri, R., et al. (2010). Comprehensive assessment of gesture production: A new test of upper limb apraxia (TULIA). *European Journal of Neurology*, 17, 59–66.

- Wang, Y. C., Kapellusch, J., & Garg, A. (2014). Important factors influencing the return to work after stroke. *Work*, 47, 553–559.
- Weiss, P. H., Kalbe, E., Kessler, J., & Fink, G. R. (2013). *Das Kölner Apraxie screening*. Göttingen: Hogrefe Verlag.
- Wirth, K., Held, A., Kalbe, E., Kessler, J., Saliger, J., Karbe, H., et al. (2016). Das revidierte Kölner Apraxie-Screening (KAS-R) als diagnostisches Verfahren für Patienten mit rechtshemipärischem Schlaganfall. [A new diagnostic tool for apraxia in patients with right-hemisphere stroke: the revised cologne apraxia screening (KAS-R)]. *Fortschritte Der Neurologie-Psychiatrie*, 84, 633–639.
- Wolf, S. L., Catlin, P. A., Ellis, M., Archer, A. L., Morgan, B., & Piacentino, A. (2001). Assessing wolf motor function test as outcome measure for research in patients after stroke. *Stroke*, 32, 1635–1639.
- Wu, A. J., Burgard, E., & Radel, J. (2014). Inpatient rehabilitation outcomes of patients with apraxia after stroke. *Topics in Stroke Rehabilitation*, 21, 211–219.
- Zago, M., & Lacquaniti, F. (2005). Cognitive, perceptual and action-oriented representations of falling objects. *Neuropsychologia*, 43, 178–188.