



Received: 02 August 2015
Accepted: 20 November 2015
Published: 11 January 2016

*Corresponding author:
Pauline E.J. Spaan, Department
of Brain and Cognition, University
of Amsterdam, Amsterdam, The
Netherlands; Department of Psychiatry
& Medical Psychology, Onze Lieve
Vrouwe Gasthuis (OLVG), Amsterdam,
The Netherlands
E-mail: paulinee.j.spaan@gmail.com

Reviewing editor:
Cornelia Duregger, Neuroconsult,
Austria

Additional information is available at
the end of the article

CLINICAL PSYCHOLOGY & NEUROPSYCHOLOGY | RESEARCH ARTICLE

Episodic and semantic memory impairments in (very) early Alzheimer's disease: The diagnostic accuracy of paired-associate learning formatsⁱ

Pauline E.J. Spaan^{1,2*}

Abstract: Paired-associate learning (PAL) paradigms measure memory processes sensitive to the medial temporal lobe, which shows atrophy in early Alzheimer's disease (AD). PAL tests have not yet been standard clinical procedure, neither are semantic memory tests. In early AD, impairments are more subtle. A literature review indicates that standard neuropsychological tests may not measure these impairments accurately. Therefore, I constructed new episodic and semantic memory tests. I investigated the diagnostic accuracy of these tests in 37 amnesic mild cognitive impairment (aMCI; of whom 21 had converted to AD at 1.3-year-follow-up), 43 early AD patients, and 80 non-demented controls. Main questions: (1) which tests best differentiate aMCI and AD from normal aging: most sensitively, most specifically?; (2) do PAL paradigms and/or semantic memory tests (fluency; naming) contribute to this differentiation? A free recall (non-PAL) test of unrelated words was most sensitive to aMCI and AD (91%), whereas a PAL-recognition-test (of semantically related word pairs of moderate association strength, including strongly related foils) was most specific (96%). Stepwise logistic regression analysis showed

ABOUT THE AUTHORS

Pauline E.J. Spaan works part-time as an assistant professor Clinical Neuropsychology at the University of Amsterdam, and part-time as a senior Clinical Neuropsychologist, conducting neuropsychological assessment and treatment, at the Department "Psychiatry and Medical Psychology" of the OLVG hospital, Amsterdam, and in her private practice.

Her research (PhD 2003) focuses on the early assessment of dementia and how various memory components may improve prediction. In particular, she is interested in the nature of semantic processing deficits in (early stage) Alzheimer's disease and how this may contribute to more reliable differential assessment.

Furthermore, she is interested in explaining patterns of age-related cognitive deficits within the normal aging spectrum: from young-old (55+) to very old age (up to 96 years old). More specifically, she studies (by means of structural equation modeling) the interplay between episodic and semantic memory components, on the one hand, and processing speed and executive functioning, on the other hand.

PUBLIC INTEREST STATEMENT

In very early Alzheimer's disease (AD), memory deficits show relatively subtle differences with memory problems seen in normal aging, non-AD-dementia or depression. Rather than administering the same tests used in clinical practice for many years, we need to develop tests that capture this subtlety. To do this, we may benefit from insights from experimental psychology, especially from research into the episodic and semantic memory deficits that are specific of early AD. I first reviewed the literature in this field to explain why I constructed my tests the way I did. This particularly involved recall and recognition of word pairs that are semantically related (e.g. "pipe" – "cigar") but not "too obvious" (e.g. "pipe" – "smoke"), because this would evoke correct answers as a result of guessing instead of actually remembering.

Results largely confirmed my theoretical expectations. This encourages further research and clinical application of these insights to improve early assessment of AD.



Pauline E.J. Spaan

that differentiation was improved by a subordinate semantic fluency test. I conclude that a combination of episodic and semantic memory components best predicts AD. Future research should focus on comparing semantic and visuospatial PAL tests.

Subjects: Aging and Health; Cognitive Neuroscience of Memory; Dementia; Dementia & Alzheimer's Disease

Keywords: paired-associate learning; recognition; semantic memory; mild cognitive impairment; Alzheimer's disease; sensitivity; specificity

1. Introduction

The neuropsychological differentiation between normal aging and the early stages of Alzheimer's disease (AD) has been investigated for many years, both from the perspective of clinical neuropsychology and from a theoretical or experimental point of view. Clinical relevance of research into the predictors of AD is obvious: early identification of AD is crucial for optimal pharmacological treatment (e.g. Flicker, 1999; Jack & Holtzman, 2013), as well as for timely provision of (psycho)social care. The first symptoms of AD are neuropsychological deficits, which may be present up to several years before a clinical diagnosis of dementia can be made (e.g. Bäckman, Jones, Berger, Laukka, & Small, 2005; Salmon, 2012; Schmand, Huizenga, & van Gool, 2010). Therefore, the search for cognitive tests that most accurately detect the underlying pattern of brain dysfunction is of great importance. This search is of theoretical importance as well: it provides knowledge about the nature of the transition from normal aging to AD by offering information on disturbed cognitive processes and neuropathology.

Spaan, Raaijmakers, and Jonker (2005) found that a verbal paired-associate learning (PAL) test, requiring cued recall of semantically related word pairs of moderate association strength, best predicted dementia (most likely AD) two years before diagnosis. This study will be discussed in more detail in Section 1.1.2. Several questions arose from this research. First of all, will the high predictive accuracy of the original PAL test be replicated when sample size is larger, and when cases are indeed in an early stage of AD, rather than dementia in general? Secondly, will the differentiation be improved by the addition of other types of episodic and/or semantic memory tests?

To investigate this, the original test battery was expanded with several newly developed episodic as well as semantic memory tests. This expansion was based on a review of the literature (Section 1.1), followed by a justification of the construction of these tests based on the literature (Section 1.2). Subsequently, in Section 1.3, I will describe the specific hypotheses of the current study.

1.1. Episodic and semantic memory measures best predicting Alzheimer's disease: a review

1.1.1. Episodic memory: the differential value of PAL tests

Various large-scale longitudinal studies have reported on the predictive value of well-known clinical tests of episodic memory (e.g. Bennett et al., 2002), particularly of measures of delayed (free) recall (e.g. Bäckman et al., 2005; Perri, Serra, Carlesimo, Caltagirone, & The Early Diagnosis Group of the Italian Interdisciplinary Network on Alzheimer's Disease, 2007). The diagnostic and prognostic accuracy of memory tests may increase if modern paradigms are used (Rentz et al., 2013). Particularly, PAL paradigms measure memory processes sensitive to the functional integrity of the medial temporal lobe (MTL) (e.g. Blackwell et al., 2004; Deweer et al., 1995; Lindeboom, Schmand, Tulner, Walstra, & Jonker, 2002; Lowndes & Savage, 2007; de Rover et al., 2011; Spaan et al., 2005; Troyer et al., 2008). This brain area is the prime location of early AD pathology (Braak & Braak, 1991). Paired associated learning entails the binding of many features of an experience to form a memory trace (e.g. Cohen et al., 1999; Eichenbaum, 1999).

Unfortunately, PAL tests have not been a traditional choice for the evaluation of memory performance of patients who are suspected to develop AD. The situation is improving, however. The visuospatial PAL test of the Cambridge Neuropsychological Test Automated Battery (CANTAB; Morris, Evenden, Sahakian, & Robbins, 1987), which requires cued recall of the locations of abstract patterns, has been found to predict AD very well (e.g. De Jager, Milwain, & Budge, 2002; Fowler, Saling, Conway, Semple, & Louis, 2002; Mitchell, Arnold, Dawson, Nestor, & Hodges, 2009). The Visual Association Test is also a PAL test, requiring cued recall of line drawings of interacting objects or animals, with high predictive validity for AD (Fuchs, Wiese, Altiner, Wollny, & Pentzek, 2012; Lindeboom et al., 2002). In addition, Rentz et al. (2013) discussed in their review several newly developed memory tests, challenging the associative binding system. These tests were found sensitive to subtle cognitive changes associated with biomarker evidence of preclinical AD. These cued recall or recognition tests included face-name pairs, shape-color combinations, pictures of common everyday objects, spatial location of dots, or word pairs.

1.1.2. Verbal paired-associate learning tests

The Wechsler memory scale (WMS; Wechsler, 1987; 1997b; Wechsler, 2009) contains a verbal PAL test. The WMS-R subtest (1987) requires cued recall of both semantically related and unrelated word pairs, whereas in the WMS-III subtest (1997b) the related pairs were omitted to increase diagnostic sensitivity. The reduced sensitivity of the related word pairs may be due to the too obvious associations (e.g. *rose-flower*). Targets with such overlearned associations to the cue should be avoided, because the answer can easily be guessed (Elias et al., 2000; Lowndes & Savage, 2007). Moreover, superordinate semantic category knowledge (e.g. *rose-flower*) is intact in AD (Rogers & Friedman, 2008), which may explain the normal performance of AD patients on the strongly related semantic word pairs of the WMS.

Nonetheless, in the WMS-IV PAL subtest (2009), four “easy” (semantically related) pairs (*door-open*; *sky-cloud*; *city-town*; *street-road*) were included again, in addition to ten “hard” (semantically unrelated) pairs (e.g. *hot-quiet*, *bag-truck*, or *day-box*). It should be noted that the easy and hard conditions differ in three respects here: degree of semantic association, degree of visual imaginability, and grammatical classification (nouns vs. adjectives or adverbs). The easy condition includes strongly semantically related pairs, 87.5% are visually imaginable words, and also 87.5% are nouns. The hard condition includes semantically unrelated pairs, only 50% are visually imaginable words and 75% are nouns. This mixture of word characteristics complicates the search for cognitive tests that most accurately detect the underlying pattern of brain dysfunction in early AD. If one condition appears to be more sensitive to early AD than the other, it is impossible to tell which of the word characteristics is responsible for the difference in sensitivity. In addition, the easy condition reintroduces the disadvantage of too obvious associations. Not surprisingly in view of the aforementioned argument, Pike et al. (2013) found that the delayed recall trial of the California Verbal Learning Test (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000) discriminated better between amnesic mild cognitive impairment (aMCI) and normal aging than the WMS-IV PAL subtest. Pike et al. nevertheless suspected that (these) word pairs are less suitable in taxing the MTL system than, for example, a visuospatial PAL test.

As was already mentioned above, but contrary to Pike et al., Spaan et al. (2005) found that a verbal PAL test best predicted the transition to dementia within two years in a heterogeneous sample of initially non-demented elderly participants ($N = 119$). At baseline, memory performance of the persons who later converted to dementia ($N = 9$; most likely AD) was characterized by a reduced ability to benefit at recall from semantic relations between words. The strength of semantic relations between cues and targets was moderate according to word association norms (de Groot, 1980; van Loon-Vervoorn & Van Bakkum, 1991). In terms of Rogers and Friedman (2008), the word pairs never had a superordinate category relationship, but were usually “attributes” (e.g. *tap-bath*) and sometimes “category coordinates” (e.g. *pipe-cigar*). In this way, correct answers as a result of overlearned associations and guessing were prevented. This verbal PAL test had better predictive validity than tests of non-PAL episodic memory (i.e. free recall and recognition of unrelated words: the 10-Word

List-Learning Test (10WLLT) and a task similar to the 10-Word-Recognition Test (10WRT), also included in the present study), category fluency (similar to the Main-Category Fluency Test (MCFT) included in the presented study), working memory (digit span and block span), and implicit memory (two priming measures derived from a word-stem completion task and a perceptual identification task, and a skill learning (mirror-reading) task). More detailed information is described in Spaan et al. (2005).

1.1.3. *The role of semantic memory*

The finding that early AD patients are less able to take advantage of semantic relations between words may be explained by poor semantic encoding (e.g. Buschke, Sliwinski, Kuslansky, & Lipton, 1997; Spaan et al., 2005). More precisely, Sailor, Bramwell, and Griesing (1998) suggested that AD patients have a deficit in the ability to evaluate semantic relations. They may no longer be able to discriminate between two related concepts, because the attribute knowledge (i.e. physical features or function) that distinguishes these two concepts has been lost. This seems consistent with the findings of Rogers and Friedman (2008) that AD patients encode information only at a rather global or superordinate level. In other words, memories become schematized, semanticized, or gist-like, with a loss of contextual details (Winocur & Moscovitch, 2011). These theoretical accounts may also explain why AD patients often produced semantically related but incorrect alternatives on a PAL test: e.g. *smoke* instead of *cigar* in response to the cue word *pipe*, or *water* instead of *bath* in response to *tap* (Spaan et al., 2005). These intrusions are overlearned associations to the cue, and consequently have higher word association frequencies than the target words (de Groot, 1980; van Loon-Vervoorn & Van Bakkum, 1991).

This raises the question whether specific tests of semantic memory, such as naming or category fluency, contribute to the differentiation between AD and normal aging, in addition to episodic memory tests. This was found in several studies in which MCI or preclinical AD patients were compared to controls (e.g. Bennett et al., 2002; Dudas, Clague, Thompson, Graham, & Hodges, 2005; Hirni, Kivisaari, Monsch, & Taylor, 2013; Lambon Ralph, Patterson, Graham, Dawson, & Hodges, 2003; Mickes et al., 2007). These results suggest dysfunction of the medial perirhinal cortex (Hirni et al., 2013) and beyond the MTL, extending to the lateral temporal lobes, in an early stage of AD (e.g. Domoto-Reilly, Sapolsky, Brickhouse, Dickerson, & for the Alzheimer's Disease Neuroimaging Initiative, 2012; Dudas et al., 2005; Hänggi, Streffer, Jäncke, & Hock, 2011; Mickes et al., 2007).

1.1.4. *Integration of episodic and semantic memory: in a PAL format*

Alternatively, one could argue that the essence of memory decline in early AD is not a matter of dysfunction of episodic and/or semantic memory, as separate or parallel impairments. Nowadays, memory theorists consider these two forms of memory (initially proposed by Tulving, 1972) as interdependent (e.g. Irish, Addis, Hodges, & Piguet, 2012). Both at encoding and at retrieval, the interdependencies between episodic and semantic memory have been demonstrated (see Greenberg & Verfaellie, 2010; for a review). For instance, new episodic learning is facilitated by an intact semantic knowledge base (e.g. Kan, Alexander, & Verfaellie, 2009), and damage to the semantic network complicates the acquisition of new episodic memories, at least in the *verbal* modality (e.g. Graham, Simons, Pratt, Patterson, & Hodges, 2000). Semantic memories are the basic material from which complex, meaningful, and detailed episodic memories are constructed, whether one is engaged in autobiographical retrieval of the past, or in constructing a plausible scenario of an event in the future (Irish & Piguet, 2013). Consequently, the degeneration of semantic memory is associated with impoverished and overgeneral episodic memory (e.g. Maguire, Kumaran, Hassabis, & Kopelman, 2010).

Also Winocur and Moscovitch (2011) emphasized the dynamic nature of memory and its underlying functional and neural interactions. Their transformation hypothesis proposes that the progression of memories from hippocampal to extra-hippocampal structures entails a loss of detailed, contextual features; as a result, memories become semanticized or gist-like, as was noted above. Memories continue to need the hippocampus to retain contextual details, which is problematic in early AD.

When we adopt these theoretical accounts, we may better understand memory problems of early AD patients as a disturbance of the *interactive* process between episodic and semantic memory, instead of (only) as isolated episodic and semantic memory deficits. A verbal PAL test is ideal to capture this interaction. Thus, a PAL test that uses semantically associated words—but *not* word pairs that represent overlearned or superordinate category relationships, or other very strong associations—may simultaneously capture three aspects of memory functioning that are impaired in early AD: (1) the episodic, MTL-or hippocampus-based memory process, (2) the semantic, extra-hippocampally based memory process, and (3) the interaction between the two.

In other words, I recommend to use tests that require paired associate learning of semantic relations that are not overlearned or superordinate. Instead, one should use word pairs that are *moderately* associated, such as category coordinates or attributes, rather than unrelated word pairs. Based on the literature review above, these task characteristics may optimize *sensitivity* to early AD.

1.1.5. *Episodic memory: the differential value of recognition formats*

Another refinement can be made in cued recall PAL tests to improve their *specificity* for MTL-related memory (encoding) deficits (e.g. Spaan, Raaijmakers, & Jonker, 2003). Because information must be encoded to be later recognized as familiar, recognition formats measure encoding rather than retrieval capacity. This results in normal or near to normal recognition performance in neurological or neuropsychiatric conditions that do *not* primarily involve MTL dysfunction (e.g. Parkinson's disease, depression) vs. reduced recognition performance in patients with MTL-related memory deficits such as AD (e.g. Lowndes & Savage, 2007) or amnesic MCI (e.g. Troyer et al., 2012). In persons suffering from non-MTL-related memory problems, free- and cued-recall usually show greater impairments than recognition memory. Recognition formats may, therefore, improve specificity of assessment, relative to a cued recall format (e.g. Bennett, Golob, Parker, & Starr, 2006; Lowndes & Savage, 2007; Lowndes et al., 2008; Pike, Rowe, Moss, & Savage, 2008; Spaan et al., 2005). Lowndes and Savage (2007) argued that a PAL task with a recognition format (i.e. a paired-associate recognition test: PART) will be both sensitive and specific of early memory change in AD. It will require structural integrity of the hippocampus (Troyer et al., 2012).

Therefore, I further refine my expectations regarding PAL tests as follows. Paired-associate recognition tests that call upon the encoding of word pairs of *moderate* semantic association strength (rather than overlearned or superordinate associations) might be particularly efficacious in the prediction of early AD. The recognition format ensures that persons without MTL damage will produce the correct answers, resulting in high specificity. I recommend the use of *distractors* that are superordinate associations or words that are strongly related to the cue. This will invite AD patients, and patients with any other MTL damage, to select the wrong answers, resulting in higher sensitivity.

1.2. *From theory to practice: efforts to improve construct validity of memory tasks*

For a valid assessment of the component processes of memory, it is crucial to administer tests that reflect these processes as purely as possible, preferably in a form that is well tolerated by elderly participants. Unfortunately, tests that are commonly used in clinical practice and in large-scale longitudinal studies do not always meet these requirements. I will discuss several disadvantages of existing tests and ways in which they may be overcome.

First of all, immediate recall measures of episodic memory such as list-learning tests are often confounded by short-term memory or attentional components as a result of recency effects. In early AD, short-term memory is unimpaired (e.g. Spaan et al., 2003). Therefore, sensitivity will improve when a test is constructed that minimizes the impact of recency effects and, instead, measures long-term memory as purely as possible. For the same reason, delayed recall measures are better predictors of AD than immediate recall measures (e.g. Bäckman et al., 2005; Perri et al., 2007). From clinical as well as empirical evidence (e.g. Bengner & Malina, 2007; Carlesimo, Marfia, Loasses, & Caltagirone, 1996), it is known that items presented last on the list are often not recalled at the delayed recall trial. However, a delayed recall trial requires a time interval that, in clinical practice, is

filled in various ways: e.g. by other tests and/or breaks, which inadvertently cause varying lengths of the interval as well as varying sorts of interference, ultimately increasing measurement error. Yet, recency effects may easily be avoided in elderly subjects, for example by a 20-s Brown–Peterson distraction task between the study and the test phase (e.g. Spaan et al., 2005). This task requires participants to count backwards by threes. This activity itself draws on short-term memory and removes the most recent items from short-term memory, leaving recall of items from long-term memory (e.g. Baddeley & Hitch, 1977). In this way, the task purity of the episodic memory test is improved, delayed recall testing is not necessary any more, and its accompanying disadvantages are avoided. These task characteristics were also already implemented in the study by Spaan et al. (2005).

Secondly, the examination of naming to confrontation (as a measure of semantic memory; which reflects an expansion compared to Spaan et al., 2005) could be improved as well. Naming tasks mostly utilize pictures (as in the Boston naming test; Kaplan, Goodglass, & Weintraub, 1983), which may introduce the unintended impact of visuoperceptual deficits (e.g. Au et al., 1995; Goulet, Ska, & Kahn, 1994). Additional administration of a task that requires naming in response to verbal definitions may contribute to the diagnostic evaluation and may also be more representative of word finding complaints in daily life (e.g. Hamberger & Seidel, 2003). Furthermore, speeded tasks that record naming latencies might be more sensitive measures of word finding capacities than unspeeded naming (e.g. Eustache, Desgranges, Jacques, & Platel, 1998).

Thirdly, category fluency tests may better reflect semantic memory functioning when more specific, subordinate semantic cues are provided (e.g. *birds*) instead of the usual superordinate cues (e.g. *animals*). In this way, a more pure semantic memory measure is obtained, on which early AD patients might perform more poorly than on superordinate semantic categories (consistent with the results of Rogers & Friedman, 2008). In addition, a subordinate category fluency test is less influenced by self-initiated retrieval or executive control processes (e.g. Mayr & Kliegl, 2000). These task characteristics also reflect an expansion compared to Spaan et al. (2005).

1.3. This study

The present study addresses the question which type of memory tests differentiate best between normal aging and aMCI or the early stage of AD. I investigated which type of *episodic* memory (free recall (i.e. non-PAL) or cued recall or recognition (PAL-based or not), with or without semantic associations; see Section 1.1) and *semantic* memory tests (naming pictures or verbal definitions, with accuracy and latencies as measures; superordinate or subordinate category fluency or phonemic fluency; see Section 1.2) differentiate best between normal aging and (pre)clinical AD. More specifically (as an expansion of the current study, in comparison with the study by Spaan et al., 2005), I examined whether the newly developed paired-associate recognition test (PART; see Section 1.1.5.), and perhaps also the additional semantic memory tests (see Section 1.1.3), contribute to the differentiation.

I expected that PAL of moderately semantically related words (see Section 1.1.4 and Spaan et al., 2005; rather than non-PAL episodic memory tests, but including a *semantic* component—rather than semantic memory tests per se (without an episodic component)) and paired-associate recognition with strongly related distractors (i.e. the PART; see Section 1.1.5) are most sensitive and most specific for early AD, respectively. This hypothesis was tested by comparing test performance of early AD and aMCI patients vs. elderly normal controls (NC).

2. Method

2.1. Participants

2.1.1. Normal controls (NC)

Participants were 171 community-dwelling volunteers between 55 and 96 years of age, recruited from different municipalities in the Netherlands through flyers and referrals from other participants.

Participants were screened for history of stroke ($N = 5$), traumatic brain injury ($N = 1$) or other neurological ($N = 1$) or psychiatric causes of cognitive dysfunction including substance abuse ($N = 3$). These subjects were excluded. Persons who used psychotropic medication ($N = 4$), did not have Dutch as native language ($N = 1$), had impaired vision that interfered with test performance ($N = 1$) or with missing values on more than two measures ($N = 4$) were also excluded. From the remaining sample of 151 NC, participants were selected that could be individually matched for age, education, and gender to individuals of the included sample of patients that is described below. Matched couples were allowed to differ a maximum of five years in age and one code of education (see Note “a” in Table 1). Prior to the test session, participants gave written informed consent. The institutional ethical review board of the department of psychology of the University of Amsterdam approved the study. Table 1 presents the characteristics of the final sample.

2.1.2. Patients

Patients were recruited from various hospitals, to which they had been referred because of memory complaints. All patients ($N = 94$) underwent comprehensive clinical, neuropsychological, neuroradiological, and laboratory assessment to objectify possible causes of memory complaints. All patients were administered a comprehensive battery of standard neuropsychological tests for the purpose of clinical assessment. The administration of these tests was *independent* of the administration of the computerized battery of episodic and semantic memory tests that was the focus of the present study. Consistent with the criteria of the above-mentioned ethical review board, all patients gave informed consent prior to the test session.

2.1.2.1. Amnestic mild cognitive impairment: For patients to be included in the aMCI group, the following diagnostic criteria (Albert et al., 2011; Petersen et al., 1999, 2001; Winblad et al., 2004) had to be met: (1) subjective memory complaints (reported by the patient and/or a significant other in a semi-structured clinical interview); (2) objective memory decline (according to the standard neuropsychological evaluation; see description below); (3) no significant decline on other cognitive functions (according to the standard neuropsychological evaluation); (4) no significant impact on carrying out daily tasks that affects independent functioning (as reported by a significant other of the patient); (5) no dementia. Exclusion criteria were a history of stroke ($N = 2$); neurological disease other than AD ($N = 1$); serious medical illness ($N = 2$) or psychiatric conditions ($N = 1$) that could cause or worsen cognitive dysfunction. Ultimately, 37 aMCI patients were included. Scores on the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) ranged between 23 and 30 ($M = 26.11$, $SD = 1.96$).

Table 1. Participant-related characteristics, per clinical group

	Clinical group							
	NC		aMCI		AD		aMCI+AD	
	N = 80		N = 37		N = 43		N = 80	
	M	SD	M	SD	M	SD	M	SD
Age (years)	74.66	7.69	73.14	7.64	75.88	6.42	74.61	7.10
Level of education ^a	3.40	1.55	3.70	1.37	2.98	1.46	3.31	1.45
Gender (% female)	73		57		63		60	
Global cognitive status (MMSE) ^{1,2,3}	28.85	1.02	26.11	1.96	23.58	2.26	24.75	2.46
Depressive symptoms (CES-D)	13.74	6.20	13.13	5.63	13.37	5.72	13.26	5.64

Notes: MMSE = Mini-Mental State Examination (Folstein et al., 1975; Maximum = 30). CES-D = Center for Epidemiologic Studies Depression Scale (Radloff, 1977). NC = normal controls. aMCI = amnestic mild cognitive impairment. AD = Alzheimer’s disease.

Paired *t* tests over: 80 (aMCI + AD) patients and 80 matched NC; 37 aMCI patients and 37 matched NC; 43 AD patients and 43 matched NC: significant difference ($p < .001$) between: ¹NC and aMCI + AD; ²NC and aMCI; ³NC and AD.

^aEducation is coded comparable to ISCED-1997 by UNESCO (2006): 0 = elementary education not completed; 1 = elementary education; 2 = less than lower vocational training; 3 = lower vocational training; 4 = intermediate vocational training, 5 = higher vocational training; 6 = university.

Memory functioning was assessed by the visual association test (Fuchs et al., 2012; Lindeboom et al., 2002) and immediate recall on the Rey complex figure test (Osterrieth, 1944; Rey, 1941), and if deemed necessary (i.e. when non-consistent results were found), by visual reproduction (WMS-III; Wechsler, 1997b), story recall (RBMT; Wilson, Cockburn, & Baddeley, 1985), and/or delayed recall of the Rey complex figure test.

Other cognitive domains (i.e. attention, working memory, executive functions, concept formation, and visuoconstruction) were assessed by digit span, symbol search test, symbol-digit coding test, similarities, and block design (Wechsler Adult Intelligence Scale–Third edition, WAIS-III; Wechsler, 1997a), copy trial of the Rey complex figure test, trail making test (Reitan, 1958), various subtests of the behavioural assessment of the dysexecutive syndrome (Wilson, Alderman, Burgess, Emslie, & Evans, 1996), Stroop color-word test (Stroop, 1935), and Raven coloured progressive matrices (Raven, 1956), and optional, letter-number sequencing (WAIS-III), Tower of London Test (Culbertson & Zillmer, 2001), Hooper visual organization test (Hooper, 1983), and/or judgement of line orientation (Benton, Hamsher, Varney, & Spreen, 1983).

Consistent with the MCI criteria (e.g. Albert et al., 2011), the aMCI patients performed deficiently on the standard memory tests (criterion 2; i.e. at least 1.5 SD (pct. < 7) below age and education corrected normative data), whereas they were not impaired on other cognitive domains (i.e. they performed on (low-)average level according to age and education corrected normative data; criterion 3; data not shown for reasons of brevity, but these can be provided on request). The patient's neurologist or geriatrician made the final diagnosis according to the MCI criteria based on the standard clinical and neuropsychological evaluation, neuro-imaging results (MRI or CT), and lab results to exclude other causes of memory decline. This assessment took place *after* the entire test administration. The neurologist or geriatrician was *blind* to performance on the episodic and semantic memory measures that were the focus of the present study.

The aMCI patients were suspected to develop AD in the upcoming years. This was supported by the observation that 21 out of 37 patients (57%)—at the time of writing, according to a review of the clinical files—had converted to AD according to consensus diagnostic criteria (McKhann et al., 1984; mean interval: 1.3 years). These patients are labeled in this study as the “preAD” (preclinical AD) patients. No systematic follow-up could be organized due to practical circumstances. Therefore, the data of this sub-sample should be viewed to be of exploratory nature. To be more specific, relevant diagnostic follow-up information was unknown for several patients because they had their medical checkup at another institution or initial assessment took place too recently ($N = 10$), or because other (medical) conditions emerged that were more urgent ($N = 3$). The remaining 3 aMCI patients did not return to the hospital anymore (for reasons unknown).

2.1.2.2. Alzheimer's disease (AD): The diagnosis of AD was made by the patient's neurologist or geriatrician according to the NINCDS-ADRDA criteria of AD (McKhann et al., 1984), based on standard neuropsychological evaluation, neuro-imaging results (MRI or CT), and lab results to exclude other causes of memory decline. Excluded were patients with a history of stroke ($N = 2$); neurological disease other than AD ($N = 1$); serious medical illness ($N = 3$) or psychiatric conditions ($N = 2$) that could cause or exacerbate cognitive dysfunction. Ultimately, 43 AD patients were included (MMSE scores ranging between 20 and 29: $M = 23.58$, $SD = 2.26$). Table 1 reports the characteristics of the final sample of 37 aMCI, 43 AD patients, and 80 matched NC.

Two paired t tests and a Pearson Chi-square test showed that the 80 NC were adequately matched to the total patient group, consisting of the 37 aMCI and the 43 AD patients, for age ($t(79) = .21$, $p = .83$), education ($t(79) = 1.04$, $p = .30$), and gender ($\chi^2(1) = 2.80$, $p = .10$). The same was true for the matching of the aMCI group and the AD group separately to their respective NC participants: age: aMCI: $t(36) = .76$, $p = .45$; AD: $t(42) = .74$, $p = .46$; education: aMCI: $t(36) = .81$, $p = .42$; AD: $t(42) = 1.27$, $p = .21$; gender: aMCI: $\chi^2(1) = .51$, $p = .48$; AD: $\chi^2(1) = .85$, $p = .36$. Furthermore, groups differed, as expected, significantly in MMSE score (aMCI: $t(36) = 8.54$, $p < .001$; AD: $t(42) = 13.03$,

$p < .001$). The degree of depressive symptoms did not differ between the groups (aMCI: $t(36) = .41$, $p = .68$; AD: $t(42) = .44$, $p = .67$).

2.2. General test procedure

Patients were tested in the hospital by a clinical neuropsychologist or a certified test technician. The NC participants were tested in their home environment by a trained neuropsychology student. The episodic and semantic memory tests were administered by means of a computer program that was operated by the experimenter. The participant only had to look at the screen, on which the stimuli were presented. The participant had to respond orally; the experimenter registered the responses. Interference between tests was prevented: e.g. there were no overlapping stimuli between tests; tests of verbal and visual nature were alternated. The sequence of administration of the tests was the same for all participants.

2.3. Test battery

Below, the episodic and semantic memory tests of the battery are briefly described. A detailed description can be found in Appendix A. Stimuli of the newly developed episodic memory tests are presented in Appendix B. See Section 1.2 for a more detailed justification of the construction of these tests, based on the literature.

2.3.1. Episodic memory

2.3.1.1. *10-Word list-learning test*. This test required free recall of 10 unrelated words in three trials. Between presentation and recall phase, a 20-s distraction task prevented recency effects.

2.3.1.2. *10-Word-recognition test*. This test involved explicit recognition (“yes” or “no”) of the words of the 10WLLT among 20 related distractors.

2.3.1.3. *Paired-associate learning test: semantic pairs*. This measure consisted of cued recall of five related word pairs, constructed in the same format as the 10WLLT.

2.3.1.4. *Paired-associate learning test: non-semantic pairs*. This measure consisted of cued recall of five unrelated word pairs.

2.3.1.5. *Paired-associate-recognition test*. This test involved explicit recognition (forced choice) of the target of the PALT, among three related distractors, in response to the cue.

2.3.2. Semantic memory

2.3.2.1. *Main-category fluency test*. This measure required the generation of exemplars from the superordinate categories of animals and products to buy in a shop within 60 s per category.

2.3.2.2. *Sub-category fluency test*. This measure required the generation of exemplars from the subcategories of birds, fishes, insects, products to buy at the greengrocery, clothes shop, and do-it-yourself shop within 30 s per subcategory.

2.3.2.3. *Letter fluency test*. This test involved generating Dutch words starting with the letters D, A, and T within 60 s per letter.

2.3.2.4. *Visual naming test: total correct*. This test involved naming 53 line drawings from the Boston naming test (Kaplan et al., 1983) appropriate for the Dutch language.

2.3.2.5. *Visual naming test: response time.* This measure represented the average response time over the correctly named items of the visual naming test.

2.3.2.6. *Verbal naming test: total correct.* This test is based on the auditory naming test (Hamberger & Seidel, 2003) and required naming of 48 descriptions of words.

2.3.2.7. *Verbal naming test: response time.* This measure represented the average response time over the correctly named descriptions of the verbal naming test.

2.4. Statistical analysis

First, paired *t* tests were performed to test for significant performance differences between the patients (the 37 aMCI and 43 AD patients separately) and their matched controls on the episodic and semantic memory measures described above. Cohen's *d* was calculated to provide information on effect size of each measure. In addition, sensitivity and specificity of classification by each measure was determined by receiver operating characteristic (ROC) analyses, providing *d'* as well as area under curve (AUC) values. These analyses were performed over the largest groups of patients (the 37 aMCI and the 43 AD patients together) and their controls.

Lastly, stepwise logistic regression analyses (using a forward likelihood ratio method) were performed to determine which combination of tests differentiated best between normal aging and early AD to answer the central research questions (see Section 1.3). In the "stepwise" method, the test measure that most successfully differentiates patients from controls is selected first. Subsequently, the test measure(s) is/are selected by the program that significantly improve(s) the prediction (or group classification), if any. Selected test measures will be presented in *decreasing* order of differential ability. Four sets of comparisons were made: (1) 43 AD patients vs. 43 matched NC participants; (2) 43 AD patients plus 37 aMCI patients vs. 80 matched NC participants; (3) 21 aMCI patients that retrospectively converted to AD according to clinical assessment at follow-up (i.e. the exploratory sub-sample of "preAD" patients) vs. 21 matched NC participants; and (4) 37 aMCI patients vs. 37 matched NC participants. Effect size of all final models was measured by Nagelkerke R^2 (Bewick, Cheek, & Ball, 2005).

3. Results

3.1. Treatment of missing data

The NC participants had no missing values; .9% of the data of the aMCI patients and 2.3% of the AD patients were missing because of fatigue. The missing values were estimated using a regression approach applying other variables that best predicted performance on the concerning measure.

3.2. Differential characteristics of each episodic and semantic memory measure

Performance on the episodic and semantic memory measures is presented in Table 2 for the samples of 37 aMCI patients, 43 AD patients, and their 80 matched NC. The estimated reliability per memory measure is also presented because I used several newly developed tasks or adapted versions of well-known tasks.

All measures had a high level of internal consistency or test-retest reliability. All measures showed a significant performance difference between the patients and their controls in the expected directions. The sample of AD patients, as well as the joined group of AD and aMCI patients, performed worse ($p < .001$) on all measures. Effect sizes were large for all episodic memory measures, both category fluency measures and the total correct score on the visual naming test (aMCI + AD vs. NC: Cohen's *d* values ranging from .82 to 1.93) and medium for the remaining measures (Cohen's *d* values ranging from .51 to .68). The aMCI patients also showed large performance differences on most measures compared to their controls ($p < .001$), with smaller differences on naming and on the letter fluency test (LFT). Effect sizes were large for all episodic memory measures, both category fluency

Table 2. Mean performance and estimated reliabilities of the episodic and semantic memory measures, per clinical group

Memory Measure	Reliability	Clinical group								
		NC			aMCI			AD		
		N = 80			N = 37			N = 43		
		M	SD	95% C.I. ^c	M	SD	95% C.I.	M	SD	95% C.I.
Episodic memory: recall measures										
10WLLT (0-30) ^{1***,2***}	.85 ^a	15.20	5.08	14.07-16.33	6.30	3.54	5.12-7.48	2.63	2.18	1.96-3.30
PALT-s (0-15) ^{1***,2***}	.80 ^a	12.11	2.47	11.56-12.66	7.08	2.75	6.16-8.00	4.00	2.63	3.19-4.81
PALT-ns (0-15) ^{1***,2***}	.83 ^a	6.00	3.46	5.23-6.77	1.57	1.82	.96-2.17	.35	.78	.11-.59
Episodic memory: recognition measures										
10WRT (0-30) ^{1***,2***}	.79 ^a	28.95	1.37	28.65-29.25	25.62	3.10	24.59-26.66	23.42	2.47	22.66-24.18
PART (0-10) ^{1***,2***}	.88 ^a	9.39	1.32	9.09-9.68	5.41	2.47	4.58-6.23	3.07	1.84	2.50-3.64
Semantic memory: fluency measures										
MCFT ^{1***,2***}	.82 ^a	38.08	8.60	36.16-39.99	29.70	5.75	27.78-31.62	22.05	5.66	20.30-23.79
SCFT ^{1***,2***}	.89 ^a	51.88	10.26	49.59-54.16	39.16	8.71	36.26-42.07	28.77	6.72	26.70-30.84
LFT ^{1*,2***}	.90 ^a	39.45	11.55	36.88-42.02	36.54	13.40	32.07-41.01	26.84	11.89	23.18-30.50
Semantic memory: naming measures										
VsNT-tc (0-53) ^{1**,2***}	.89 ^a	47.44	4.19	46.50-48.37	44.46	6.46	42.30-46.61	35.72	8.41	33.13-38.31
VsNT-rt (s) ^{1*,2***}	.84 ^{b**}	3.08	.81	2.90-3.26	3.20	.57	3.02-3.39	3.93	.83	3.67-4.18
VbNT-tc (0-48) ^{1*,2***}	.89 ^a	45.14	2.76	44.52-45.75	43.92	3.52	42.74-45.09	38.03	6.15	36.14-39.92
VbNT-rt (s) ^{1***,2***}	.81 ^{b**}	4.52	1.11	4.27-4.76	5.05	.88	4.76-5.35	6.09	1.20	5.72-6.46

Notes: Possible ranges per memory measure are presented between parentheses. 10WLLT = 10-word list-learning test. PALT-s = paired-associate learning test: semantic pairs. PALT-ns = paired-associate learning test: non-semantic pairs. 10WRT = 10-word-recognition test. PART = paired-associate-recognition test. MCFT = main-category fluency test. SCFT = sub-category fluency test. LFT = letter fluency test. VsNT-tc = visual naming test: total correct. VsNT-rt = visual naming test: response time. VbNT-tc = verbal naming test: total correct. VbNT-rt = verbal naming test: response time. NC = normal controls. aMCI = amnesic mild cognitive impairment. AD = Alzheimer's disease.

^aCronbach's alpha determined over a sample (N = 341), also including excluded participants.

^bBecause Cronbach's alpha could not be calculated for these measures, test-retest reliability (including level of significance) was calculated instead, derived from 24 NC that were retested two years later.

^c95% confidence interval for the mean.

Paired t tests over NC, aMCI, and AD groups.

*Significant difference at $p < .05$.

**Significant difference at $p < .01$.

***Significant difference at $p < .001$ between: ¹NC and aMCI; ²NC and AD.

measures and the response time on the visual naming test (Cohen's d values ranging from .84 to 1.62) and small for the remaining measures (Cohen's d values ranging from .36 to .42).

Table 3 presents the classification characteristics of each memory measure for the aMCI and AD patients together vs. their controls. The paired-associate recognition test (PART) had the highest differential ability. Consistent with my hypothesis described in Sections 1.1.5 and 1.3, this was mainly

Table 3. Classification characteristics of specific episodic and semantic memory measures in differentiating between the group of 80 patients assessed with Alzheimer’s disease (N = 43) or amnesic mild cognitive impairment (N = 37) and 80 cognitively healthy elderly controls (matched for age, education, and gender)

Memory measure	Correctly classified (%)	Sensitivity (%)	Specificity (%)	d'	AUC (SE)
Episodic memory: recall measures					
10WLLT	89.4	91.3	87.5	2.5	.96 (.02)
PALT-s	85.6	90.0	81.2	2.2	.94 (.02)
PALT-ns	82.5	85.0	80.0	1.9	.91 (.02)
Episodic memory: recognition measures					
10WRT	86.3	85.0	87.5	2.2	.91 (.03)
PART	91.3	86.3	96.3	2.9	.96 (.02)
Semantic memory: fluency measures					
MCFT	76.3	76.3	76.3	1.4	.88 (.03)
SCFT	83.8	86.3	81.3	2.0	.91 (.02)
LFT	65.0	62.5	67.5	.8	.69 (.04)
Semantic memory: naming measures					
VsNT-tc	75.0	66.3	83.8	1.4	.80 (.04)
VsNT-rt	63.8	60.0	67.5	.7	.70 (.04)
VbNT-tc	66.3	58.8	73.8%	.9	.76 (.04)
VbNT-rt	70.0	66.3	73.8%	1.1	.78 (.04)
MMSE ^a	85.6	83.8	87.5%	2.1	.93 (.02)

Notes: 10WLLT = 10-word list-learning test. PALT-s = paired-associate learning test: semantic pairs. PALT-ns = paired-associate learning test: non-semantic pairs. 10WRT = 10-word-recognition test. PART = paired-associate-recognition test. MCFT = main-category fluency test. SCFT = sub-category fluency test. LFT = letter fluency test. VsNT-tc = visual naming test: total correct. VsNT-rt = visual naming test: response time. VbNT-tc = verbal naming test: total correct. VbNT-rt = verbal naming test: response time. MMSE = Mini-Mental State Examination. AUC = area under curve (ROC curve).

^aDetermined for reasons of comparison.

due to the very high level of specificity, whereas sensitivity was still satisfactory. Only the 10-word list-learning test (10WLLT) and the number of semantic pairs recalled on the paired-associate learning test (PALT-s) had higher sensitivity (i.e. the latter (not the first) also consistent with my hypothesis described in Sections 1.1.4 and 1.3). The 10WLLT also had high specificity. The “Paired-Associate Learning Test: non-semantic pairs” (PALT-ns) differentiated less well. Performance of aMCI and AD patients on this test was close to floor level, whereas performance of NC participants was rather variable (see Table 2; this seems consistent with the rationale described in Sections 1.1.2 and 1.1.4). The other recognition test (10WRT) had a sensitivity level similar to the PART, but its specificity was lower.

Concerning the semantic memory measures, only the sub-category fluency test (SCFT) showed adequate levels of sensitivity and specificity. This seems consistent with the rationale described in Sections 1.1.3 and 1.2. The MCFT and particularly the LFT had much lower levels of sensitivity and specificity. Sensitivity of the naming measures was also low, although specificity of the total correct score on the visual naming test (VsNT-tc) was reasonable.

For reasons of comparison, the classification characteristics of the MMSE were also determined (see Table 3). Note that the MMSE is an important variable in clinical dementia assessment, and group classification is partly dependent on the MMSE. Nevertheless, differential ability of the MMSE does not (clearly) exceed differential ability of the PART, 10WLLT, PALT-s, and SCFT. Particularly, sensitivity of the 10WLLT and specificity of the PART are evidently better than of the MMSE.

3.3. Memory measures best predicting Alzheimer's disease

Finally, I investigated which combination of tests most accurately classified the early AD patients among non-demented controls. Table 4 shows the best differentiating tests, in *decreasing order* of differential ability, according to “stepwise” logistic regression analyses. These analyses were performed over various subsamples of aMCI and AD patients and their matched controls, to investigate whether the best differentiating tests varied according to severity of symptoms (i.e. the proportion of clinically diagnosed AD patients).

In comparisons of AD patients vs. controls and AD plus aMCI patients vs. controls (comparisons 1 and 2; Table 4), the PART was selected first by the “stepwise” method. Thus, it may be concluded that the PART differentiated best when the patient sample included a relatively high proportion of clinically diagnosed AD patients. Also in aMCI patients that were retrospectively diagnosed with AD (comparison 3: the preAD subgroup), the PART contributes significantly to the differentiation. This seems consistent with its high specificity (see Section 3.2), and with the second part of my hypothesis. In comparison 4, the PART was *not* selected as one of the best differentiating tests. A closer look at the results showed that the PART *did* contribute significantly to aMCI classification accuracy, but at step 1 the 10WLLT (score 39.55, $p < .001$, vs. 38.52, $p < .001$) contributed slightly better, and at step 2 the PALT-s (score 13.94, $p < .001$, vs. 10.05, $p = .002$) contributed slightly better (see Table Note “b”). Therefore, I conclude that the PART is certainly not irrelevant, but slightly less successful to

Table 4. Combination of episodic and semantic memory measures best differentiating between each of four samples of patients with (preclinical) Alzheimer's disease or amnesic mild cognitive impairment and cognitively healthy elderly controls (matched for age, education, and gender) and their classification accuracy and the effect size of the final predictive model (measured by Nagelkerke R^2)

Best differentiating memory measures ^a	Correctly classified (%)	Sensitivity	Specificity	d'	AUC (SE)	Nagelkerke R^2
(1) 43 AD vs. 43 NC						
PART + 10WLLT + SCFT	100	43/43 (100%)	43/43 (100%)	∞	1 (0)	1.00
(2) (37 aMCI + 43 AD) vs. 80 NC						
PART + SCFT + PALT-s + 10WLLT	95.0	76/80 (95.0%)	76/80 (95.0%)	3.3	.99 (.007)	.89
(3) 21 preAD vs. 21 NC						
10WLLT + PALT-s + PART	100	21/21 (100%)	21/21 (100%)	∞	1 (0)	1.00
(4) 37 aMCI vs. 37 NC						
10WLLT + PALT-s ^b + SCFT	98.6	37/37 (100%)	36/37 (97.3%)	∞	.98 (.022)	.87
Specific combinations of measures ^c	Correctly classified	Sensitivity	Specificity	d'	AUC (SE)	Nagelkerke R^2
(37 aMCI + 43 AD) vs. 80 NC						
PART + 10WLLT	92.5	74/80 (92.5%)	74/80 (92.5%)	2.9	.98 (.009)	.84
PART + PALT-s	91.9	74/80 (92.5%)	73/80 (91.3%)	2.8	.97 (.013)	.81
PART + PALT-s + PALT-ns	92.5	73/80 (91.3%)	75/80 (93.8%)	2.9	.97 (.012)	.82

Notes: AD = Alzheimer's disease. NC = normal controls. aMCI = amnesic mild cognitive impairment. preAD = preclinical AD (i.e. aMCI patients that were retrospectively formally diagnosed with AD). PART = paired-associate-recognition test. 10WLLT = 10-word list-learning test. SCFT = sub-category fluency test. PALT-s = paired-associate learning test: semantic pairs. AUC = area under curve (of the ROC curve).

^aIn *decreasing order* of differential ability, according to *stepwise* logistic regression analysis.

^b10WLLT and PALT-s were of slightly higher significance to enter the equation at steps 1 and 2, respectively, than PART.

^cSpecific combination of measures, *entered together* (not *stepwise*) in the logistic regression analysis, on an exploratory basis.

detect AD at a very early or preclinical stage. Although it is not certain that all aMCI patients were in fact in a preclinical stage of AD (i.e. which is certain in the preAD subgroup).

Compared to the PART, the 10WLLT, seemed better able to detect AD at the earliest (preclinical) stage (comparisons 3 and 4; see Table 4). In comparisons 1 and 2, it was not selected first, but still contributed significantly to the classification, whereas in comparisons 3 and 4 it was the best predictor of group classification. Furthermore, a semantic memory measure—the SCFT—significantly improved the differentiation, in three out of four comparisons (except for comparison 3, focused on the 21 preAD patients). Thus, the differentiating value of the SCFT does not seem to be affected by the proportion of clinically diagnosed AD patients. The same was true for the PALT-s.

Classification accuracy was very high in all four comparisons. In addition, effect size of all final models, as measured by Nagelkerke R^2 , was large ($>.80$). This indicates that the explanatory variables are highly useful in detecting early AD (e.g. Bewick et al., 2005). Consistently, the set of best differentiating tests is similar over all four comparisons.

On an exploratory basis, I conducted a few additional analyses over the largest groups of participants (comparison 2), entering the best differentiating tests together (instead of using the stepwise method). As is shown in the lower part of Table 4, the combination of the two best differentiating measures (PART and 10WLLT) showed an excellent accuracy of classification. The same level of classification success was achieved by the combination of the PART, PALT-s, and PALT-ns (and nearly the same level when the PALT-ns was left out).

4. Discussion

In the present study, the diagnostic accuracy of a computerized battery of episodic and semantic memory tests was investigated in a group of amnesic MCI and early AD patients, relative to cognitively healthy elderly controls. The tests were constructed to measure specific memory components as purely as possible by minimizing the impact of short-term memory on episodic memory and of executive control processes on semantic memory. In addition, the tests were selected to create a varied battery involving: free recall; cued recall of semantically unrelated and related words (avoiding overlearned or superordinate associations); recognition (in forced choice and yes/no-format); superordinate and subordinate category fluency and phonemic fluency; accuracy and speed of naming of pictures and verbal descriptions of words.

4.1. Episodic and semantic memory impairments in (pre)clinical Alzheimer's disease

The main question was which type of memory tests differentiated best between normal aging and early or preclinical AD. In a previous study, Spaan et al. (2005) found that dementia was best predicted, two years before diagnosis but in a very small sample ($N = 9$), by reduced benefit of semantic relations (of moderate association strength) in a paired-associate learning test—the PALT-s measure in the current study. In the current study, I investigated whether the differentiating value of a (semantic) PAL paradigm could be replicated in a much larger sample of patients, who were actually diagnosed with AD, or who (probably) were in a preclinical stage (the aMCI patients). More precisely, I examined whether the newly developed paired-associate recognition test (PART) and perhaps also semantic memory tests improved the differentiation, particularly concerning specificity of classification.

The results showed that the PART (indeed) and the 10-word list-learning test (10WLLT; contrary to the hypothesis) were, respectively, most specific and most sensitive to AD. However, the combination of the PART and the PALT-s was nearly as successful. From a practical point of view, the latter combination is more time-efficient than the first, because administration of the PART requires the preceding administration of the PALT, during which the word-pairs are studied. In any case, the PALT and the PART classified early AD more accurately than the MMSE; despite the fact that the MMSE partly determined group classification, whereas the PALT and the PART did not.

4.1.1. *The diagnostic accuracy of PAL formats*

Consistent with the literature reviewed in Section 1.1., PAL paradigms indeed appear to be very sensitive to MTL-related memory deficits characteristic of AD (e.g. Lowndes & Savage, 2007; Troyer et al., 2008). The 10WLLT proved to be highly sensitive as well, but it was less specific, presumably because it does not involve the binding of to-be-learned stimuli as in PAL tests. Instead, the 10WLLT demands self-initiated effortful search strategies that may be independent of MTL functioning. These processes may be sensitive to disturbances within, for example, the prefrontal cortex that also occur in normal aging, especially at very old age (e.g. Crawford, Bryan, Luszcz, Obonsawin, & Stewart, 2000; Dempster, 1992). This may explain the relatively low specificity of the 10WLLT.

The results indicate that episodic memory tests that call upon efficient semantic association of words are most accurate in the prediction or classification of AD. This may also explain why the PART obtained a higher specificity than the 10 word-recognition test (10WRT). The latter test less easily evokes semantic associations (and thereby false recognitions, common in early AD), as these were not induced during the study phase in the 10WLLT. Furthermore, cued recall of semantically *unrelated* word pairs (PALT-ns) was less accurate in classifying AD. This is in contrast with some studies, which intentionally tested recall of semantically unrelated words in order to increase sensitivity (e.g. Lowndes et al., 2008; Wechsler, 1997b). However, the type of semantic association between words in a PAL paradigm may be crucial: whereas overlearned or superordinate semantic knowledge is intact in AD, adequate encoding of moderately associated words (representing an “attribute” or “category coordinate” relationship) causes problems (e.g. Rogers & Friedman, 2008; Sailor et al., 1998; Spaan et al., 2005).

4.1.2. *The role of semantic memory*

From the findings described in the previous section, it may be concluded that poor performance on episodic memory tests that require a sufficiently deep and detailed, and a less gist-like level of semantic processing is most predictive of early AD (see also Winocur & Moscovitch, 2011). In addition, it is probably not a coincidence that the SCFT contributed to the differentiation. This is a more purely semantic memory test than standard superordinate category fluency tests (i.e. the MCFT). The MCFT did not contribute in the multivariate models (Table 4) and it had a clearly lower differential ability than most of the other tests (Table 3). The diagnostic value of the SCFT is consistent with studies reporting additional differential ability of certain semantic memory tests (e.g. Dudas et al., 2005; Hirni et al., 2013; Mickes et al., 2007).

Thus, a combination of episodic and semantic memory components seems to detect early AD best. This is consistent with recent theoretical views on memory fractionation stating that episodic and semantic memory are interdependent (Greenberg & Verfaellie, 2010) or dynamically interacting (Winocur & Moscovitch, 2011). I argue that the PALT-s and the PART are memory tests in which this interdependence is crucial. Both tests require adequate binding of contextual (i.e. episodic) information and information stored in semantic memory (e.g. Reder, Park, & Kieffaber, 2009). The additional value of the PALT-s and the PART, above traditional memory tests including most PAL paradigms, is owed to the selection of the stimulus materials: word pairs of moderate semantic association strength in the first test, and distractor words that were strongly related to the cue word in the second test. These task characteristics require a high quality of binding between episodic and semantic information to prevent mistakes. This probably renders these tests more appropriate than conventional tests, such as the 10WLLT, PALT-ns, and 10WRT (or *visual* PAL tests reviewed in Section 1.1.1, although this was not investigated in this study), to detect the memory problems of early AD but not of normal aging.

4.2. *Strengths and limitations of the present study and future research directions*

A strength of the present study is my attempt to improve construct validity and increase task purity of the tests I administered. Most important in this respect, and innovative compared to previous studies, was the critical selection of word pairs and distractors of the PALT-s and the PART, and the avoidance in both recall tests (10WLLT and PALT) of the unintended impact of short-term memory

as a result of recency and sequential learning effects. In addition, the fact that the test battery was computerized probably reduced measurement error, thus increasing test reliability (Snyder et al., 2011). Also Rentz et al. (2013) recommended computerized testing.

However, it should be noted that this study was not specifically aimed at systematically investigating which specific task conditions (mentioned above) better differentiated between normal aging and early AD than others. This could still be explored in future research. Although these kind of issues were not investigated within an experimental research design, the relatively better differential ability of some tasks compared to other tasks is still clinically and theoretically interesting.

In addition, it would be interesting to directly compare the differentiating value of the PAL paradigms that were found useful in the current study (i.e. PART and PALT-s, involving carefully chosen *semantic* memory components), with *visuospatial* PAL paradigms (as in the CANTAB; Morris et al., 1987), or *visual* PAL paradigms, as in the visual-association test (Lindeboom et al., 2002). It may be noted that the PALT predicted demented cases more accurately than the visual-association test in the Spaan et al. (2005) study. Nonetheless, a new direct comparison in a larger study would still be interesting, especially when also a visuospatial PAL test is included. In any case, one should prevent circularity between independent (diagnosis) research variables and dependent (test performance) research variables.

Limitations of my study may be the moderate sample size and the lack of a systematic follow-up assessment procedure in the aMCI subsample. However, the effect sizes of all differentiating tests and of all final predictive logistic regression models were large. In addition, the set of best differentiating tests was highly similar over four performed analyses or (sub)samples. Although I have no (systematic) follow-up information of all MCI patients, recent literature (Albert et al., 2011; Dubois et al., 2007) shows that it is likely that this group contains a large proportion of very early AD patients. This was also indicated by the exploratory subsample of 21 preAD patients.

5. General conclusions and recommendations for clinical practice

My results suggest that a carefully chosen combination of episodic and semantic memory components, integrated within single test paradigms, is able to almost perfectly separate amnesic MCI and early AD from normal aging. This is in accordance with recent theoretical views on memory organization that emphasize the interdependence and dynamic interaction between episodic and semantic memory (Greenberg & Verfaellie, 2010; Winocur & Moscovitch, 2011). A verbal PAL paradigm—particularly using a recognition format—was found to be most adequate. The type of semantic associations incorporated in a PAL paradigm seems crucial for obtaining high sensitivity. Moderately associated words should be selected rather than more strongly related words that represent overlearned or superordinate semantic associations. Whether this type of (semantic) PAL paradigm predicts AD better (or not) than, for example, visuospatial PAL paradigms (e.g. De Jager et al., 2002; Fowler et al., 2002; Mitchell et al., 2009) still has to be investigated in future research, also taking into account the heterogeneity of Alzheimer's disease.

Experimental-neuropsychological studies have reported numerous interesting findings that could be useful for clinical practice. I hope to have shown with this study that the search for memory measures that best detect early AD benefits from joining knowledge and methods from clinical neuropsychology and experimental memory psychology (e.g. Lowndes & Savage, 2007; Spaan et al., 2003).

Acknowledgements

The author gratefully acknowledges Prof. Dr Ben Schmand, Prof. Dr Jeroen Raaijmakers, and Prof. Dr Jaap Murre for helpful comments on an earlier draft of this manuscript, and Dr Verena Schmittmann for methodological advice. In addition, I wish to thank Suk Fang Hau, Claudia Ootjers, Renske Nekkers, Antoinet Vierich, Sanne Bentvelzen, and Trudy van der Geest for data collection. I also thank

the departments of Geriatrics, Neurology, and Medical Psychology of the "Groene Hart Ziekenhuis", Gouda, the "Kennemer Gasthuis", Haarlem, the "Medisch Spectrum Twente", Enschede, and the "Streekziekenhuis Koningin Beatrix", Winterswijk.

Funding

The author received no direct funding for this research.

Competing interests

The author declares no competing interests.

Author details

Pauline E.J. Spaan^{1,2}

E-mail: pauline.j.spaan@gmail.com

ORCID ID: <http://orcid.org/0000-0001-5858-2419>

¹ Department of Brain and Cognition, University of Amsterdam, Amsterdam, The Netherlands.

² Department of Psychiatry & Medical Psychology, Onze Lieve Vrouwe Gasthuis (OLVG), Amsterdam, The Netherlands.

Citation information

Cite this article as: Episodic and semantic memory impairments in (very) early Alzheimer's disease: The diagnostic accuracy of paired-associate learning formats, Pauline E.J. Spaan, *Cogent Psychology* (2016), 3: 1125076.

Note

i The research described in this research article was presented both at the 2010 Mid-Year Meeting of the International Neuropsychological Society, Krakow, Poland, July 2010, and at the Second Meeting of the Federation of European Societies of Neuropsychology, Amsterdam, the Netherlands, September 2010.

References

- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., ... Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7, 270–279.
- Au, R., Joung, P., Nicholas, M., Obler, L. K., Kass, R., & Albert, M. L. (1995). Naming ability across the adult life span. *Aging and Cognition*, 2, 300–311.
- Baayen, R.H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database (CD-ROM)* (Linguistic Data Consortium). Philadelphia, PA: University of Pennsylvania.
- Bäckman, L., Jones, S., Berger, A. K., Laukka, E. J., & Small, B. J. (2005). Cognitive impairment in preclinical Alzheimer's disease: A meta-analysis. *Neuropsychology*, 19, 520–531. <http://dx.doi.org/10.1037/0894-4105.19.4.520>
- Baddeley, A. D., & Hitch, G. J. (1977). Recency re-examined. In S. Dornic (Ed.), *Attention and performance* (Vol. 6, pp. 646–667). Hillsdale, NJ: Erlbaum.
- Bengner, T., & Malina, T. (2007). Dissociation of short- and long-term face memory: Evidence from long-term recency effects in temporal lobe epilepsy. *Brain and Cognition*, 64, 189–200. <http://dx.doi.org/10.1016/j.bandc.2007.02.006>
- Bennett, D. A., Wilson, R. S., Schneider, J. A., Evans, D. A., Beckett, L. A., Aggarwal, N. T., ... Bach, J. (2002). Natural history of mild cognitive impairment in older persons. *Neurology*, 59, 198–205. <http://dx.doi.org/10.1212/WNL.59.2.198>
- Bennett, I. J., Golob, E. J., Parker, E. S., & Starr, A. (2006). Memory evaluation in mild cognitive impairment using recall and recognition tests. *Journal of Clinical and Experimental Neuropsychology*, 28, 1408–1422. <http://dx.doi.org/10.1080/13803390500409583>
- Benton, A. L., & Hamscher, K. (1976). *Multilingual aphasia examination*. Iowa City: University of Iowa.
- Benton, A. L., Hamscher, K., Varney, N. R., & Spreen, O. (1983). *Contributions to neuropsychological assessment*. New York, NY: Oxford University Press.
- Bewick, V., Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression. *Critical Care*, 9, 112–118. <http://dx.doi.org/10.1186/cc3045>
- Blackwell, A. D., Sahakian, B. J., Vesey, R., Semple, J. M., Robbins, T. W., & Hodges, J. R. (2004). Detecting dementia: Novel neuropsychological markers of preclinical Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, 17, 42–48. <http://dx.doi.org/10.1159/000074081>
- Braak, H., & Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*, 82, 239–259. <http://dx.doi.org/10.1007/BF00308809>
- Buschke, H., Sliwinski, M. J., Kuslansky, G., & Lipton, R. B. (1997). Diagnosis of early dementia by the double memory test: Encoding specificity improves diagnostic sensitivity and specificity. *Neurology*, 48, 989–996. <http://dx.doi.org/10.1212/WNL.48.4.989>
- Carlesimo, G. A., Marfia, G. A., Loasses, A., & Caltagirone, C. (1996). Recency effect in anterograde amnesia: Evidence for distinct memory stores underlying enhanced retrieval of terminal items in immediate and delayed recall paradigms. *Neuropsychologia*, 34, 177–184. [http://dx.doi.org/10.1016/0028-3932\(95\)00100-X](http://dx.doi.org/10.1016/0028-3932(95)00100-X)
- Cohen, N., Ryan, J., Hunt, C., Romine, L., Wszalek, T., & Nash, C. (1999). Hippocampal system and declarative (relational) memory: Summarizing the data from functional neuroimaging studies. *Hippocampus*, 9, 83–98. [http://dx.doi.org/10.1002/\(ISSN\)1098-1063](http://dx.doi.org/10.1002/(ISSN)1098-1063)
- Crawford, J. R., Bryan, J., Luszcz, M. A., Obonsawin, M. C., & Stewart, L. (2000). The executive decline hypothesis of cognitive aging: Do executive deficits qualify as differential deficits and do they mediate age-related memory decline? *Aging, Neuropsychology, and Cognition*, 7, 9–31. <http://dx.doi.org/10.1076/anec.7.1.9.806>
- Culbertson, W., & Zillmer, E. (2001). *Tower of London Drexel University (TOL DX)*. Chicago, IL: Multi-Health Systems.
- De Jager, C. A., Milwain, E., & Budge, M. (2002). Early detection of isolated memory deficits in the elderly: The need for more sensitive neuropsychological tests. *Psychological Medicine*, 32, 483–491.
- Delis, D. C., Kramer, J., Kaplan, E., & Ober, B. (2000). *CVLT-II: California verbal learning test second edition adult version*. Antonio, TX: Psychological Corporation.
- Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review*, 12, 45–75. [http://dx.doi.org/10.1016/0273-2297\(92\)90003-K](http://dx.doi.org/10.1016/0273-2297(92)90003-K)
- de Rover, M., Pironti, V. A., McCabe, J. A., Acosta-Cabrero, J., Arana, F., Morein-Zamir, S., ... Sahakian, B. J. (2011). Hippocampal dysfunction in patients with mild cognitive impairment: A functional neuroimaging study of a visuospatial paired associates learning task. *Neuropsychologia*, 49, 2060–2070. <http://dx.doi.org/10.1016/j.neuropsychologia.2011.03.037>
- Deweere, B., Lehericy, S., Pillon, B., Baulac, M., Chiras, J., Marsault, C., ... Dubois, B. (1995). Memory disorders in probable Alzheimer's disease: The role of hippocampal atrophy as shown with MRI. *Journal of Neurology, Neurosurgery and Psychiatry*, 58, 590–597. <http://dx.doi.org/10.1136/jnnp.58.5.590>
- Domoto-Reilly, K., Sapolsky, D., Brickhouse, M., Dickerson, B. C., & for the Alzheimer's Disease Neuroimaging Initiative (2012). Naming impairment in Alzheimer's disease is associated with left anterior temporal lobe atrophy. *NeuroImage*, 63, 348–355. <http://dx.doi.org/10.1016/j.neuroimage.2012.06.018>
- Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., ... Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *The Lancet Neurology*, 6, 734–746. [http://dx.doi.org/10.1016/S1474-4422\(07\)70178-3](http://dx.doi.org/10.1016/S1474-4422(07)70178-3)
- Dudas, R. B., Clague, F., Thompson, S. A., Graham, K. S., & Hodges, J. R. (2005). Episodic and semantic memory

- in mild cognitive impairment. *Neuropsychologia*, 43, 1266–1276.
<http://dx.doi.org/10.1016/j.neuropsychologia.2004.12.005>
- Eichenbaum, H. (1999). The hippocampus and mechanisms of declarative memory. *Behavioural Brain Research*, 103, 123–133.
[http://dx.doi.org/10.1016/S0166-4328\(99\)00044-3](http://dx.doi.org/10.1016/S0166-4328(99)00044-3)
- Elias, M. F., Beiser, A., Wolf, P. A., Au, R., White, R. F., & D'Agostino, R. B. (2000). The preclinical phase of Alzheimer disease. *Archives of Neurology*, 57, 808–813.
<http://dx.doi.org/10.1001/archneur.57.6.808>
- Eustache, F., Desgranges, B., Jacques, V., & Platel, H. (1998). Preservation of the attribute knowledge of concepts in normal aging groups. *Perceptual and Motor Skills*, 87, 1155–1162.
<http://dx.doi.org/10.2466/pms.1998.87.3f.1155>
- Flicker, L. (1999). Acetylcholinesterase inhibitors for Alzheimer's disease. *British Medical Journal*, 318, 615–616.
<http://dx.doi.org/10.1136/bmj.318.7184.615>
- Folstein, M. F., Folstein, S. F., & McHugh, P. R. (1975). "Mental state". *Journal of Psychiatric Research*, 12, 189–198.
[http://dx.doi.org/10.1016/0022-3956\(75\)90026-6](http://dx.doi.org/10.1016/0022-3956(75)90026-6)
- Fowler, K. S., Saling, M. M., Conway, E. L., Semple, J. M., & Louis, W. J. (2002). Paired associate performance in the early detection of DAT. *Journal of the International Neuropsychological Society*, 8, 58–71.
<http://dx.doi.org/10.1017/S1355617702811067>
- Fuchs, A., Wiese, B., Altiner, A., Wollny, A., & Pentzek, M. (2012). Cued recall and other cognitive tasks to facilitate dementia recognition in primary care. *Journal of the American Geriatrics Society*, 60, 130–135.
<http://dx.doi.org/10.1111/jgs.2012.60.issue-1>
- Goulet, P., Ska, B., & Kahn, H. J. (1994). Is there a decline in picture naming with advancing age? *Journal of Speech Language and Hearing Research*, 37, 629–644.
<http://dx.doi.org/10.1044/jshr.3703.629>
- Graham, K. S., Simons, J. S., Pratt, K. H., Patterson, K., & Hodges, J. R. (2000). Insights from semantic dementia on the relationship between episodic and semantic memory. *Neuropsychologia*, 38, 313–324.
[http://dx.doi.org/10.1016/S0028-3932\(99\)00073-1](http://dx.doi.org/10.1016/S0028-3932(99)00073-1)
- Greenberg, D. L., & Verfaellie, M. (2010). Interdependence of episodic and semantic memory: Evidence from neuropsychology. *Journal of the International Neuropsychological Society*, 16, 748–753.
<http://dx.doi.org/10.1017/S1355617710000676>
- de Groot, A. M. B. (1980). *Mondelinge woordassociatienormen: 100 woordassociaties op 460 Nederlandse zelfstandige naamwoorden* [Oral word association norms: 100 word associations to 460 Dutch nouns]. Lisse: Swets & Zeitlinger.
- Hamberger, M. J., & Seidel, W. T. (2003). Auditory and visual naming tests: Normative and patient data for accuracy, response time, and tip-of-the-tongue. *Journal of the International Neuropsychological Society*, 9, 479–489.
- Hänggi, J., Streffer, J., Jäncke, L., & Hock, C. (2011). Volumes of lateral temporal and parietal structures distinguish between healthy aging, mild cognitive impairment, and Alzheimer's disease. *Journal of Alzheimer's Disease*, 26, 719–734.
- Hirni, D. I., Kivisaari, S. L., Monsch, A. U., & Taylor, K. I. (2013). Distinct neuroanatomical bases of episodic and semantic memory performance in Alzheimer's disease. *Neuropsychologia*, 51, 930–937.
<http://dx.doi.org/10.1016/j.neuropsychologia.2013.01.013>
- Hooper, H. E. (1983). *Hooper visual organization test (VOT)*. Los Angeles, CA: Western Psychological Services.
- Irish, M., Addis, D. R., Hodges, J. R., & Piguet, O. (2012). Considering the role of semantic memory in episodic future thinking: Evidence from semantic dementia. *Brain*, 135, 2178–2191.
<http://dx.doi.org/10.1093/brain/aws119>
- Irish, M., & Piguet, O. (2013). The pivotal role of semantic memory in remembering the past and imagining the future. *Frontiers in Behavioral Neuroscience*, 7, doi: <http://dx.doi.org/10.3389/fnbeh.2013.00027>
- Jack, C. R., & Holtzman, D. M. (2013). Biomarker modeling of Alzheimer's disease. *Neuron*, 80, 1347–1358.
<http://dx.doi.org/10.1016/j.neuron.2013.12.003>
- Kan, I. P., Alexander, M. P., & Verfaellie, M. (2009). Contribution of prior semantic knowledge to new episodic learning in amnesia. *Journal of Cognitive Neuroscience*, 21, 938–944.
<http://dx.doi.org/10.1162/jocn.2009.21066>
- Kaplan, E. F., Goodglass, H., & Weintraub, S. (1983). *The Boston naming test* (2nd ed.). Philadelphia, PA: Lea & Febiger.
- Lambon Ralph, M. A., Patterson, K., Graham, N., Dawson, K., & Hodges, J. R. (2003). Homogeneity and heterogeneity in mild cognitive impairment and Alzheimer's disease: A cross-sectional and longitudinal study of 55 cases. *Brain*, 126, 2350–2362.
<http://dx.doi.org/10.1093/brain/awg236>
- Lindeboom, J., Schmand, B., Tulner, L., Walstra, G., & Jonker, C. (2002). Visual association test to detect early dementia of the Alzheimer type. *Journal of Neurology, Neurosurgery and Psychiatry*, 73, 126–133.
<http://dx.doi.org/10.1136/jnnp.73.2.126>
- van Loon-Vervoorn, W. A., & Van Bakkum, I. J. (1991). *Woordassociatie lexicon: Gebaseerd op 1299 woorden en 100 proefpersonen* [Word association lexicon: Based on 1299 words and 100 subjects]. Amsterdam/Lisse: Swets & Zeitlinger.
- Lowndes, G., & Savage, G. (2007). Early detection of memory impairment in Alzheimer's disease: A neurocognitive perspective on assessment. *Neuropsychology Review*, 17, 193–202.
<http://dx.doi.org/10.1007/s11065-007-9032-z>
- Lowndes, G. J., Saling, M. M., Ames, D., Chiu, E., Gonzalez, L. M., & Savage, G. R. (2008). Recall and recognition of verbal paired associates in early Alzheimer's disease. *Journal of the International Neuropsychological Society*, 14, 591–600.
- Maguire, E. A., Kumaran, D., Hassabis, D., & Kopelman, M. D. (2010). Autobiographical memory in semantic dementia: A longitudinal fMRI study. *Neuropsychologia*, 48, 123–136.
<http://dx.doi.org/10.1016/j.neuropsychologia.2009.08.020>
- Marien, P., Mampaey, E., Vervaeke, A., Saerens, J., & De Deyn, P. P. (1998). Normative data for the Boston naming test in native Dutch-speaking Belgian elderly. *Brain and Language*, 65, 447–467.
<http://dx.doi.org/10.1006/brln.1998.2000>
- Mayr, U., & Kliegl, R. (2000). Complex semantic processing in old age: Does it stay or does it go? *Psychology and Aging*, 15, 29–43.
<http://dx.doi.org/10.1037/0882-7974.15.1.29>
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34, 939–939.
<http://dx.doi.org/10.1212/WNL.34.7.939>
- Mickes, L., Wixted, J. T., Fennema-Notestine, C., Galasko, D., Bondi, M. W., Thal, L. J., & Salmon, D. P. (2007). Progressive impairment on neuropsychological tasks in a longitudinal study of preclinical Alzheimer's disease. *Neuropsychology*, 21, 696–705.
<http://dx.doi.org/10.1037/0894-4105.21.6.696>
- Mitchell, J., Arnold, R., Dawson, K., Nestor, P. J., & Hodges, J. R. (2009). Outcome in subgroups of mild cognitive impairment (MCI) is highly predictable using a simple algorithm. *Journal of Neurology*, 256, 1500–1509.
<http://dx.doi.org/10.1007/s00415-009-5152-0>

- Morris, R., Evenden, J., Sahakian, B., & Robbins, T. (1987). Computer aided assessment of dementia: Comparative studies of neuropsychological deficits in Alzheimer type dementia and Parkinson's disease. In S. Stahl, S. Iversen, & E. Goodman (Eds.), *Cognitive neurochemistry* (pp. 21–36). Oxford: Oxford University Press.
- Osterrieth, P. A. (1944). Le test de copie d'une figure complexe: Contribution à l'étude de l'aperception et de la mémoire [The copy test of a complex figure: Contribution to the study of perception and memory]. *Archives de Psychologie*, 30, 205–353.
- Perri, R. Serra, L., Carlesimo, G. A., Caltagirone, C., & The Early Diagnosis Group of the Italian Interdisciplinary Network on Alzheimer's Disease. (2007). Amnesic mild cognitive impairment: Difference of memory profile in subjects who converted or did not convert to Alzheimer's disease. *Neuropsychology*, 21, 549–558. <http://dx.doi.org/10.1037/0894-4105.21.5.549>
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment. *Archives of Neurology*, 56, 303–308. <http://dx.doi.org/10.1001/archneur.56.3.303>
- Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., ... Winblad, B. (2001). Current concepts in mild cognitive impairment. *Archives of Neurology*, 58, 1985–1992. <http://dx.doi.org/10.1001/archneur.58.12.1985>
- Pike, K. E., Kinsella, G. J., Ong, B., Mullaly, E., Rand, E., Storey, E., ... Parsons, S. (2013). Is the WMS-IV Verbal paired associates as effective as other memory tasks in discriminating amnesic mild cognitive impairment from normal aging? *The Clinical Neuropsychologist*, 27, 908–923. <http://dx.doi.org/10.1080/13854046.2013.809149>
- Pike, K. E., Rowe, C. C., Moss, S. A., & Savage, G. (2008). Memory profiling with paired associate learning in Alzheimer's disease, mild cognitive impairment, and healthy aging. *Neuropsychology*, 22, 718–728. <http://dx.doi.org/10.1037/a0013050>
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385–401. <http://dx.doi.org/10.1177/014662167700100306>
- Raven, J. C. (1956). *Coloured progressive matrices*. Oxford: Oxford Psychologists Press.
- Reder, L. M., Park, H., & Kieffaber, P. D. (2009). Memory systems do not divide on consciousness: Reinterpreting memory in terms of activation and binding. *Psychological Bulletin*, 135, 23–49. <http://dx.doi.org/10.1037/a0013974>
- Reitan, R. M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Perceptual and Motor Skills*, 8, 271–276. <http://dx.doi.org/10.2466/pms.1958.8.3.271>
- Rentz, D. M., Parra Rodriguez, M. A., Amariglio, R., Stern, Y., Sperling, R., & Ferris, S. (2013). Promising developments in neuropsychological approaches for the detection of preclinical Alzheimer's disease: A selective review. *Alzheimer's Research and Therapy*, 5, 58. <http://dx.doi.org/10.1186/alzrt222>
- Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique [The psychological examination of cases of traumatic encephalopathy]. *Archives de Psychologie*, 28, 286–340.
- Rogers, S. L., & Friedman, R. B. (2008). The underlying mechanisms of semantic memory loss in Alzheimer's disease and semantic dementia. *Neuropsychologia*, 46, 12–21. <http://dx.doi.org/10.1016/j.neuropsychologia.2007.08.010>
- Sailor, K. M., Bramwell, A., & Griesing, T. A. (1998). Evidence for an impaired ability to determine semantic relations in Alzheimer's disease patients. *Neuropsychology*, 12, 555–564. <http://dx.doi.org/10.1037/0894-4105.12.4.555>
- Salmon, D. P. (2012). Neuropsychological features of mild cognitive impairment and preclinical Alzheimer's disease. In M. C. Pardon & M. W. Bondi (Eds.), *Behavioral neurobiology of aging* (pp. 187–212), *Current topics in behavioral neurosciences*, 10. Berlin: Springer Berlin Heidelberg.
- Schmand, B., Groenink, S. C., & den Dungen, M. (2008). Letterfluency: Psychometrische eigenschappen en Nederlandse normen. *Tijdschrift voor Gerontologie en Geriatrie*, 39, 64–74. <http://dx.doi.org/10.1007/BF03078128>
- Schmand, B., Huizenga, H. M., & van Gool, W. A. (2010). Meta-analysis of CSF and MRI biomarkers for detecting preclinical Alzheimer's disease. *Psychological Medicine*, 40, 135–145. <http://dx.doi.org/10.1017/S0033291709991516>
- Snyder, P. J., Jackson, C. E., Petersen, R. C., Khachaturian, A. S., Kaye, J., Albert, M. S., & Weintraub, S. (2011). Assessment of cognition in mild cognitive impairment: A comparative study. *Alzheimer's & Dementia*, 7, 338–355.
- Spaan, P. E. J., Raaijmakers, J. G. W., & Jonker, C. (2003). Alzheimer's disease versus normal ageing: A review of the efficiency of clinical and experimental memory measures. *Journal of Clinical and Experimental Neuropsychology (Neuropsychology, Development and Cognition: Section A)*, 25, 216–233. <http://dx.doi.org/10.1076/jcen.25.2.216.13638>
- Spaan, P. E. J., Raaijmakers, J. G. W., & Jonker, C. (2005). Early assessment of dementia: The contribution of different memory components. *Neuropsychology*, 19, 629–640. <http://dx.doi.org/10.1037/0894-4105.19.5.629>
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662. <http://dx.doi.org/10.1037/h0054651>
- Troyer, A. K., Murphy, K. J., Anderson, N. D., Craik, F. I. M., Moscovitch, M., Maione, A., & Gao, F. (2012). Associative recognition in mild cognitive impairment: Relationship to hippocampal volume and apolipoprotein E. *Neuropsychologia*, 50, 3721–3728. <http://dx.doi.org/10.1016/j.neuropsychologia.2012.10.018>
- Troyer, A. K., Murphy, K. J., Anderson, N. D., Hayman Abello, B. A., Craik, F. I. M., & Moscovitch, M. (2008). Item and associative memory in amnesic mild cognitive impairment: Performance on standardized memory tests. *Neuropsychology*, 22, 10–16. <http://dx.doi.org/10.1037/0894-4105.22.1.10>
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). New York, NY: Academic Press.
- UNESCO. (2006). *International Standard Classification of Education: ISCED-1997*. Retrieved from <http://www.uis.unesco.org/Library/Documents/isced97-en.pdf>
- van den Dungen, M., & Groenink, S. (2004). *Normeringsonderzoek voor de Nederlandse letterfluency* [Standardization study on Dutch letter fluency] (Unpublished master's thesis). University of Amsterdam, Amsterdam.
- Wechsler, D. (1987). *Wechsler memory scale-revised*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1997a). *Wechsler adult intelligence scale* (3rd ed.). San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1997b). *Wechsler memory scale* (3rd ed.). San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2009). *Wechsler memory scale* (4th ed.). San Antonio, TX: Pearson.
- Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H. E., & Evans, J. J. (1996). *Behavioural assessment of dysexecutive syndrome*. Bury St. Edmunds: Thames Valley Test.
- Wilson, B. A., Cockburn, J., & Baddeley, A. D. (1985). *The Rivermead behavioural memory test*. Bury St. Edmunds: Thames Valley Test.
- Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L.-O., et al. (2004). Mild cognitive impairment—Beyond controversies, towards a consensus: Report

of the International Working Group on mild cognitive impairment. *Journal of Internal Medicine*, 256, 240–246.
<http://dx.doi.org/10.1111/jim.2004.256.issue-3>

Winocur, G., & Moscovitch, M. (2011). Memory transformation and systems consolidation. *Journal of the International Neuropsychological Society*, 17, 1–15.

Appendix A

Detailed description of the measures of the test battery

A.1. Measurement variables

Below, the measures of the test battery are briefly described, ordered by the dependent, latent cognitive construct of which they were an observed variable.

A.1.1. Episodic memory measures

See Appendix B for the stimuli (and their order) in each of the episodic memory tests. The 10WLLT, 10WRT, PALT-s, PALT-ns, and the “animal”-category of the MCFT were also administered in Spaan et al. (2005). In addition, all measures were included in a larger study on normal aging (Spaan, 2015).

A.1.1.1. 10-word list-learning test (10WLLT). This test required free recall of 10 semantically unrelated words (concrete imaginable nouns), in three trials. Each word was presented on the screen for 2 s, with an interval of .5 s. Between presentation and recall phase (to prevent recency effects), the participant performed a 20-s Brown–Peterson distraction task, requiring to count backwards from a three-digit number that was presented on the screen. In each trial, the same words were presented but in a different sequence to prevent sequential learning effects. The score was the total number of words reproduced over three trials (possible range of scores: 0–30).

A.1.1.2. 10-word-recognition test (10WRT). This test involved explicit recognition (by answering “yes” or “no”) of the words of the 10-word list-learning test from a list that also included 20 semantically related distractor words. The distractors were not previously presented and were matched to the targets for word length. The 10-word-recognition test was administered immediately after the 10-word list-learning test. The score was the sum of true positive and true negative answers (range: 0–30).

A.1.1.3. Paired-associate learning test: semantic pairs (PALT-s). This measure consisted of cued recall of five semantically related word pairs, constructed in the same format as in the 10-word list-learning test. Three trials were given. During the study phase, each pair was presented on the screen for 3 s, with an interval of .5 s. The participant was not alerted to the semantic association. Semantic associations were derived from Dutch word association norms (de Groot, 1980; van Loon-Vervoorn & Van Bakkum, 1991). The target word of each pair was moderately semantically related to the cue word, to prevent correct answers as a result of free association or guessing (see detailed description below, under “Paired-Associate-Recognition Test”). Words were matched for word length and lexical frequency (Baayen, Piepenbrock, & Van Rijn, 1993). Between the study and the recall phase, the participant performed the same 20-s distraction task as in the “10-Word List-Learning Test”. During the recall phase, the participant was asked to name the target word in response to each cue word; the cues were consecutively presented on the screen in an order different from the presentation order during the study phase. In case of incorrect answers, the experimenter mentioned that the answer was not correct (to prevent misunderstandings about the test instructions or error-full learning), but without providing the correct answer. In each trial, the same word pairs were presented but in a different sequence to prevent sequential learning effects. The score was the sum of pairs reproduced over three trials (range: 0–15).

A.1.1.4. Paired-associate learning test: non-semantic pairs (PALT-ns). This measure consisted of cued recall of five semantically unrelated word pairs, as part of the paired-associate learning test described above. The score was the sum of pairs reproduced over three trials (range: 0–15).

A.1.1.5. Paired-associate-recognition test (PART). This test involved explicit recognition (forced choice) of the target words of the paired-associate learning Test in response to the presented cue words. In each trial, three distractor words were simultaneously presented on the screen together with the target and the cue (see Figures 1a and 1b). Semantic associations were derived from the same Dutch word association norms (de Groot, 1980; van Loon-Vervoorn & Van Bakkum, 1991). Words were matched for word length. The score was the sum of correct answers (range: 0–10).

A.1.2. Semantic memory measures

A.1.2.1. Main-category fluency test. This measure required the participant to generate as many exemplars that belonged to the categories of “animals” and “products to buy in a shop” as he/she could think of within 60 s per category. The score was the sum of correct and unique answers over both categories.

A.1.2.2. Sub-category fluency test. As part of the category fluency test mentioned above, the participant was required to generate as many exemplars that belonged to the subcategories of “birds”,

Figure 1a. Visualization of an item of the semantic condition of the paired-associate-recognition test (PART).

Notes: pipe–cigar. The first distractor (“smoke”) was more strongly semantically related to the cue (“pipe”), compared to the association between the target (“cigar”) and the cue, whereas the second distractor (“tobacco”) was of similar association strength; the third distractor was semantically unrelated to the cue (“soup”).

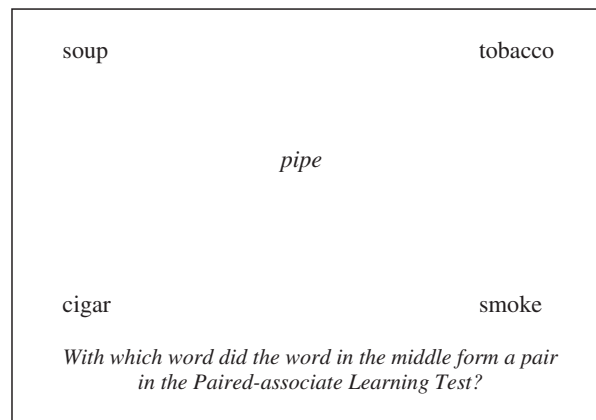
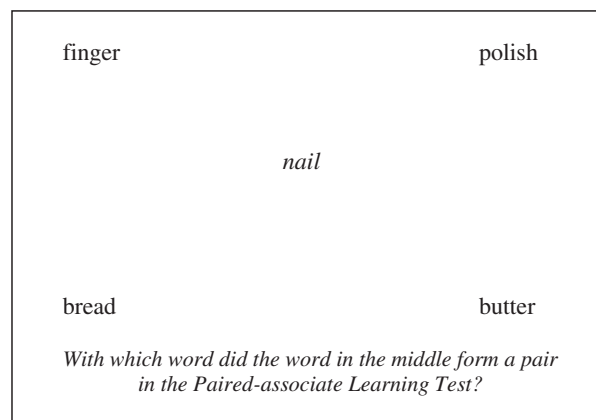


Figure 1b. Visualization of an item of the non-semantic condition of the paired-associate-recognition test (PART).

Notes: nail–butter. The first distractor (“finger”) was strongly semantically related to the cue (“nail”), whereas the second distractor (“bread”) was strongly semantically related to the target (“butter”); the third distractor (“polish”) was moderately related to the cue.



“fishes”, “insects”, “products to buy at the greengrocery”, “products to buy in a clothes shop”, and “products to buy in a do-it-yourself shop” as one could think of within 30 s per subcategory. By providing more specific semantic cues (as compared to the superordinate cues of the MCFT), a more pure semantic memory measure is obtained that requires less self-initiated retrieval or executive control processes. The score was the sum of correct and unique answers over six subcategories.

A.1.2.3. Letter fluency test. This test involved generating as many Dutch words starting with the letters D, A, and T as one could think of within 60 s per letter (Schmand, Groenink, & den Dungen, 2008; i.e. the Dutch version of the Controlled Oral Word Association Task; Benton & Hamsher, 1976). The score was the sum of correct and unique answers over three letters.

A.1.2.4. Visual naming test: total correct (VsNT-tc). This test involved naming line drawings of common objects and animals, presented on the screen. Material was taken from the Boston naming test (Kaplan et al., 1983) and consisted of 53 of the 60 original stimuli appropriate for the Dutch language (which were appropriate for the Dutch language; Marien, Mampaey, Vervaeke, Saerens, & De Deyn, 1998). The score was the sum of spontaneously and correctly named pictures within 20 s per item (range: 0–53).

A.1.2.5. Visual naming test: response time (VsNT-rt). This measure represented the average response time (in s) determined over the spontaneously, correctly named pictures of the visual naming test, described above. Response times were recorded by the experimenter (by a mouse click), as soon as the participant correctly named a picture.

A.1.2.6. Verbal naming test: total correct (VbNT-tc). This test is a computerized variant of the auditory naming test developed by Hamberger and Seidel (2003). Instead of the experimenter reading aloud the descriptions of the words (which the participant had to name), the descriptions were presented on the screen (to minimize working memory load and to overcome hearing difficulties). Stimuli used by Hamberger and Seidel were translated into Dutch by van den Dungen and Groenink (2004), which lead to a few alterations in order to create 48 descriptions appropriate for the Dutch language. The score was the sum of spontaneously and correctly named descriptions within 20 s per item (range: 0–48).

A.1.2.7. Verbal naming test: response time (VbNT-rt). This measure represented the average response time (in s) determined over the spontaneously, correctly named descriptions of the verbal naming test, described above. The registered response time reflected the time between presentation of the description on the screen and the moment that the participant gave a correct answer (within 20 s). Response times were recorded by the experimenter by a mouse click.

A.2. References

- Baayen, R.H., Piepenbrock, R., & Van Rijn, H. (1993). The CELEX lexical database (CD-ROM). Linguistic Data Consortium. Philadelphia, PA: University of Pennsylvania.
- Benton, A.L., & Hamsher, K. (1976). Multilingual aphasia examination. Iowa City: University of Iowa.
- Dungen, van den, M., & Groenink, S. (2004). Normeringsonderzoek voor de Nederlandse letterfluency [Standardization study on Dutch letter fluency] (Unpublished master's thesis). University of Amsterdam, Amsterdam.
- Groot, de, A. M. B. (1980). Mondelinge woordassociatienormen: 100 woordassociaties op 460 Nederlandse zelfstandige naamwoorden [Oral word association norms: 100 word associations to 460 Dutch nouns]. Lisse: Swets & Zeitlinger.
- Hamberger, M. J., & Seidel, W. T. (2003). Auditory and visual naming tests: Normative and patient data for accuracy, response time, and tip-of-the-tongue. *Journal of the International Neuropsychological Society*, 9, 479–489.
- Kaplan, E. F., Goodglass, H., & Weintraub, S. (1983). The Boston naming test (2nd ed.). Philadelphia, PA: Lea & Febiger.

Loon-Vervoorn, van, W. A., & Van Bakkum, I. J. (1991). *Woordassociatie lexicon: Gebaseerd op 1299 woorden en 100 proefpersonen* [Word association lexicon: Based on 1299 words and 100 subjects]. Amsterdam/Lisse: Swets & Zeitlinger.

Mariën, P., Mampaey, E., Vervaeke, A., Saerens, J., & De Deyn, P.P. (1998). Normative data for the Boston Naming Test in native Dutch-speaking Belgian elderly. *Brain and Language*, 65, 447–467.

Schmand, B., Groenink, S. C., & Van den Dungen, M. (2008). Letterfluency: Psychometrische eigenschappen en Nederlandse normen [Letterfluency: Psychometric properties and Dutch normative data]. *Tijdschrift voor Gerontologie en Geriatrie*, 39, 64–76.

Spaan, P. E. J. (2015). Episodic and semantic memory functioning in very old age: Explanations from executive functioning and processing speed theories. *Cogent Psychology*, 2: 1109782.

Spaan, P. E. J., Raaijmakers, J. G. W., & Jonker, C. (2005). Early assessment of dementia: The contribution of different memory components. *Neuropsychology*, 19, 629–640.

Appendix B: Stimuli in each of the episodic memory tests

See Appendix A for the detailed (procedural) description of the respective tests.

Table 5a. Order of items (in Dutch), per trial, in the “10-Word List-Learning Test” (see Table 5b for the English translation of the words)

Trial 1	Trial 2	Trial 3
eland	loods	viltstift
kermis	wond	kermis
polder	eland	wond
viltstift	polder	ballon
ballon	ballon	polder
wond	viltstift	loods
afval	kermis	litteken
zakdoek	litteken	zakdoek
loods	afval	afval
litteken	zakdoek	eland

Table 5b. Order of items (in Dutch) in the “10-Word-Recognition Test” (including English translation)

Trial	Answer	English translation
puin	No	Debris
eland	Yes	Elk
viltstift	Yes	Felt-pen
papier	No	Paper
afval	Yes	Waste (matter)
vallei	No	Valley
pleister	No	Bandage
kraam	No	Booth
ballon	Yes	Balloon
heuvel	No	Hill
ree	No	Deer
snee	No	Cut
lampion	No	Lampion
loods	Yes	Shed
balpen	No	ball Point

(Continued)

Table 5b. (Continued)

Trial	Answer	English translation
kist	No	Box/case
wond	Yes	Wound
slinger	No	Garland
zakdoek	Yes	Handkerchief
circus	No	Circus
polder	Yes	Polder
zwellling	No	Swelling
hinde	No	Hind/doe
breuk	No	Fracture
litteken	Yes	Scar
kluis	No	Safe
handdoek	No	Towel
vulpen	No	Fountain pen
vuilnis	No	Rubbish
kermis	Yes	(Fun) fair

Table 5c. Order of items (in Dutch), per trial, in the study phase and in the recall phase, in the “Paired-Associate Learning Test”; English translation included

Trial 1: pairs	Translation	Trial 1: cue	Trial 2: pairs	Trial 2: cue	Trial 3: pairs	Trial 3: cue
mouw-trui	<i>sleeve-sweater</i>	rat	nagel-boter	mouw	streep-groente	schilder
salon-druif	<i>salon-grape</i>	nagel	salon-druif	varken	kraan-bad	mouw
storm-appel	<i>storm-apple</i>	pijp	varken-modder	rat	nagel-boter	streep
kraan-bad	<i>(water) tap-bath</i>	salon	rat-knecht	kraan	schilder-verf	storm
varken-modder	<i>pig-mud</i>	storm	storm-appel	nagel	salon-druif	varken
streep-groente	<i>stripe-vegetable</i>	schilder	pijp-sigaar	schilder	storm-appel	salon
pijp-sigaar	<i>pipe-cigar</i>	varken	streep-groente	pijp	mouw-trui	rat
nagel-boter	<i>nail-butter</i>	kraan	mouw-trui	salon	varken-modder	pijp
rat-knecht	<i>rat-servant</i>	mouw	kraan-bad	streep	pijp-sigaar	nagel
schilder-verf	<i>painter-paint</i>	streep	schilder-verf	storm	rat-knecht	kraan

Table 5d. Order of items (in Dutch) in the “Paired-Associate-Recognition Test” (the cue word is presented first; see also Figures 1a and 1b); English translation included

Trial 1	Translation
pijp-soep-tabak-sigaar-rook	<i>pipe-soup-tobacco-cigar-smoke</i>
schilder-huis-verf-kwast-aardappel	<i>painter-house-paint-brush-potato</i>
streep-sla-lijn-potlood-groente	<i>stripe-lettuce-line-pencil-vegetable</i>
nagel-vinger-lak-brood-boter	<i>nail-finger-polish-bread-butter</i>

(Continued)

Table 5d. (Continued)

Trial 1	Translation
mouw-jas-cent-hemd-trui	<i>sleeve-coat-cent-tank top/shirt-sweater</i>
varken-kaars-modder-koe-vlees	<i>pig-candle-mud-cow-meat</i>
rat-knecht-slaaf-staart-muis	<i>rat-servant-slave-tail-mouse</i>
storm-appel-regen-wind-peer	<i>storm-apple-rain-wind-pear</i>
salon-kamer-druif-vrucht-tafel	<i>salon-room-grape-fruit-table</i>
kraan-water-tap-bad-haard	<i>(water) tap-water-tap/bar-bath-fire place</i>



© 2016 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.



Cogent Psychology (ISSN: 2331-1908) is published by Cogent OA, part of Taylor & Francis Group.

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com

