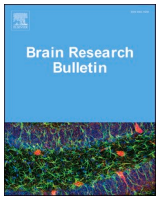
Brain Research Bulletin 229 (2025) 111439

Contents lists available at ScienceDirect

Brain Research Bulletin

journal homepage: www.elsevier.com/locate/brainresbull



Research report

Semantic distances of WAIS Similarities word pairs in non-demented adults: An item-level index of semantic memory granularity

Matteo De Marco [a,*], Satyam Chauhan [a], Martina Bocchetta [a,b], Annalena Venneri [a,c]

[a] Centre for Cognitive and Clinical Neuroscience, Department of Psychology, College of Health, Medicine and Life Sciences, Brunel University of London, Uxbridge, UK
[b] Dementia Research Centre, Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, University College London, London, UK
[c] Department of Medicine and Surgery, University of Parma, Italy

ARTICLE INFO

Keywords:
Alzheimer's disease
Mild cognitive impairment
Transentorhinal
Perirhinal
Semantic

ABSTRACT

Innovative approaches to test scoring can help neuropsychologists detect subtle semantic memory alterations. We focussed on the Wechsler Adult Intelligence Scale Similarities (SIM) test and calculated an item-level score based on the 'Leacock & Chodorow '(LCH) semantic distance expressed by each SIM item. We hypothesised that LCH would predict 1) performance on standard semantic memory tests; 2) an Alzheimer aetiology; 3) transentorhinal grey-matter integrity. Six hundred sixty-nine non-demented participants completed a neuropsychological battery inclusive of SIM and consolidated tests of semantic memory and executive functioning. Hierarchical linear regressions were designed to test the association between LCH and semantic memory performance after controlling for major confounders. A hierarchical logistic regression was then designed to test the association between LCH and underlying aetiology (Alzheimer/cerebrovascular) in a mild cognitive impairment sub-cohort. Finally, we tested the association between LCH and both whole-brain grey-matter density and transentorhinal thickness using voxel-based-morphometry and region-of-interest models. LCH predicted semantic memory performance but not on a test significantly supported by executive resources. LCH also predicted clinical aetiology and grey-matter density in the transentorhinal cortex and in other regions involved in linguistic-semantic processing. No significant association was found with regional thickness. *Post-hoc* LCH scoring in 89 people with dementia revealed the presence of a gradient of diagnostic severity, i.e., healthy adults < mild cognitive impairment < dementia. Item-level scores of SIM performance are associated with neurocognitive constituents of semantic memory. LCH is a valuable construct that could help clinicians detect semantic memory decline in ageing adults with suspected neurodegeneration.

"Pour l'esprit de finesse, nous proposons de donner à définir, à indiquer les ressemblances et les différences entre deux ou plusieurs synonymes, par exemple entre 'bonté', 'tendresse' et 'amabilité'; le sujet devrait écrire quelles sont les différences et les ressemblances de ces expressions."

Alfred Binet and Victor Henri, 1895

1. Introduction

Having reached its fifth edition as of September 2024, the Wechsler Adult Intelligence Scale (WAIS) is one of the most widely-used instruments to assess cognition (Wechsler, 2008). One of its sections, the Similarities sub-test (SIM), is a verbal task in which the testee is presented with word pairs and, for each pair, as per the manufacturer's instructions, "*is asked to identify the qualitative relationship between the two words*".

Presented as a test of *concept formation*, *abstract thinking* and *verbal reasoning* (see Box 1 for the definition of these notions provided by the American Psychological Association), SIM performance relies on two of the six cognitive domains that are recognised by the Diagnostic and Statistical Manual of Mental Disorders (DSM)-5 (Sachdev et al., 2014): 'learning and memory' (i.e., in its 'semantic and autobiographical long-term memory' subdomain) and 'executive function'. The former is associated with the retrieval of semantic knowledge that is conveyed by the words, while the latter is associated with the "mental manipulations"

* Correspondence to: Department of Psychology, College of Health, Medicine and Life Sciences, Brunel University of London, Kingston Lane, Uxbridge, Middlesex UB8 3PH, UK.
E-mail address: matteo.demarco@brunel.ac.uk (M. De Marco).

https://doi.org/10.1016/j.brainresbull.2025.111439
Received 26 January 2025; Received in revised form 4 June 2025; Accepted 22 June 2025
Available online 23 June 2025

that are demanded by abstracting and reasoning. The executive aspect of SIM is also highlighted by the inclusion of this task in the Frontal Assessment Battery as a measure sensitive to frontal lobe damage (Dubois et al., 2000) and by evidence collected in neurological settings whereby statistical effects in SIM performance were interpreted as driven by executing functioning (Giovannetti et al., 2001; Pérez-Cordón et al., 2020; Stokholm et al., 2006; Zalonis et al., 2012).

Thanks to its sensitivity to semantic processes, SIM is of particular relevance to the study of preclinical Alzheimer's disease (AD) detection. Initiatives that recruited cohorts of cognitively healthy adults and followed them up over time while monitoring their neurological trajectories indicate that objective alterations to semantic processing are detectable 6–12 years before a clinical diagnosis of AD is made (Amieva et al., 2008; Hirni et al., 2016; Payton et al., 2020). This is because, within the network of regions known to contribute to semantic processing, e.g., left temporal pole (Herlin et al., 2021) and inferior parietal lobule (Binder and Desai, 2011), semantic memory retrieval is supported by the transentorhinal cortex (Clarke and Tyler, 2014; Mishkin et al., 1997), and this is the earliest region to harbour hyperphosphorylated tau pathology (i.e., neurofibrillary tangles and neuropil threads) in AD at Braak Stage I (Braak and Braak, 1995; Braak et al., 2006), at a stage when no objective episodic memory deficits are yet reported. The mild cognitive impairment (MCI) phase (i.e., the phase when objective deficits emerge), in fact, typically corresponds to Braak Stages II to IV (Markesbery, 2010). These are the neuropathological grounds by which semantic memory test scores that are only minimally influenced by executive functioning would be extremely valuable to detect neuropsychological changes due to Braak Stage I hyperphosphorylated tau pathology. Based on this, while the combined influence of semantic and executive abilities described above may affect SIM performance when the test is scored according to its standard procedures, an alternative scoring approach based on the *semantic difficulty* of each item could provide additional information about the performance that is less influenced by executive functioning. On this note, "item-level" scoring approaches have been proposed in relation to a number of cognitive clinical tests of memory and language, such as Category Fluency (Forbes-McKay et al., 2005; Vonk et al., 2019), Boston Naming (De Marco et al., 2023a), or Prose Memory (Mueller et al., 2023). With a specific focus on SIM performance, an item-level scoring approach would grade each item according to its semantic difficulty, on the basis that semantically more distant words would require more difficult processing.

In this study, we quantified the semantic distance between the two words of each of the 19 WAIS III SIM pairs, and we used this information to characterise the granularity of semantic processing of a large transdiagnostic cohort of cognitively normal adults and individuals with MCI. Our overarching hypothesis posited that the qualitative item-level scores derived from this procedure would be a significant predictor of key neurocognitive and clinical constituents of semantic memory processing, after controlling for standard quantitative SIM scores. In line with this hypothesis, we formulated three predictions. Firstly, qualitative item-level scores would predict performance on established tests of semantic memory, after controlling for standard quantitative SIM scores (*Prediction 1*). Secondly, we expected that a sub-sample of individuals with MCI of clinically-established AD aetiology, due to their underlying semantic difficulties, would show poorer item-level scores than MCI individuals of clinically-established cerebrovascular aetiology (*Prediction 2*). Finally, we predicted that item-level scores would be associated with the integrity of the transentorhinal cortex, as assessed with brain MRI measures (*Prediction 3*).

2. Methods

2.1. Participants

A cohort of 732 individuals not fulfilling a diagnosis of dementia were recruited at the tertiary-care memory clinic of Sheffield Teaching Hospitals (Sheffield, UK) between 2011 and 2020, to be enrolled in clinical research protocols. All individuals were seen by a clinical neuropsychologist and completed an extensive battery of cognitive tests (listed in Table 1). Of these 732 individuals, 270 were invited to the clinic as healthy adults, while the remaining 462 had been referred to a neuropsychological assessment by a neurologist, for suspected cognitive decline. In line with the (now historical) clinical recommendations by Petersen and colleagues (Petersen et al., 2018), diagnoses of MCI were not systematically supported by a biomarker profile. Although MCI is a diagnostic entity characterised by homogeneous clinical severity and neurobiological heterogeneity (Winblad et al., 2004), the goal of *Prediction 1* was, in fact, to define a statistical link within a set of demographic and neuropsychological variables, regardless of the underlying mechanisms inducing cognitive impairment. The study was approved by the West of Scotland Regional Ethics Committee 5, Ref. No.: 19/WS/0177. Each participant (including those described in **Section 3.4**) provided their written consent prior to taking part in the study.

2.2. Standard scoring of SIM performance

The 19 SIM items were initially scored based on the standard WAIS guidelines. Items 1–5 were scored 0 or 1, while items 6–19 were given a score of 0, 1 or 2, depending on the accuracy of the response (Wechsler, 2008). As a result, standard scores ranged between 0 and 33. Data from 63 participants were excluded due to non-completion ($n = 31$) or low (≤ 9, i.e., a number of correct responses that was insufficient to calculate adequate item-level scores) SIM scores ($n = 32$), leaving a final sample of $n = 669$. Of these, 259 were healthy controls, 114 were diagnosed with amnestic MCI, 229 were diagnosed with non-amnestic MCI, and 67 did not show any objective cognitive impairment and were classified as having subjective cognitive complaints. A flowchart illustrating these numbers is included in Fig. 1. Diagnostic status was binarised to establish the presence/absence of cognitive impairment (i.e., "MCI" and "controls"), with amnestic and non-amnestic MCI individuals and, similarly, healthy volunteers and individuals diagnosed with subjective cognitive complaints being grouped together.

2.3. Item-level scoring of SIM performance

The WordNet (version 3.0) initiative (https://wordnet.princeton.edu/) is a freely available resource that represents all English

Box 1
The three abilities on which SIM performance relies, as defined by the American Psychological Association.

- Abstract thinking: "*thinking characterized by the use of general ideas or concepts*".
- Concept formation: "*the process by which a person abstracts a common idea from one or more particular examples and learns the defining features or combination of features that are characteristic of a class (e.g., those describing a bird) or that are necessary and sufficient to identify members of a class of objects, relations, or actions (e.g., the concepts* triangle, above, *or* move*)*".
- Reasoning: "*thinking in which logical processes of an inductive or deductive character are used to draw conclusions from facts or premises*".

Table 1
Characterisation of the extended cohort split by diagnostic status.

	Entire Cohort	MCI	Controls	Sub-Sample Size (MCI-Controls)
Demographic/Clinical Variable				
Age (years)	58.22 (17.33)	63.41 (10.39)	52.74 (21.10)	343–324
Education (years)	13.59 (3.33)	12.40 (2.82)	14.84 (3.37)	342–324
Sex (F/M)	332/337	157/186	175/151	343–326
Mini Mental State Examination	27.65 (1.82)	26.72 (1.76)	28.63 (1.29)	343–326
Confrontational Naming	19.21 (1.12)	19.12 (1.27)	19.32 (0.93)	341–323
Pyramids & Palm Trees Test	50.44 (1.90)	50.16 (2.28)	50.73 (1.32)	339–322
Category Fluency Test	45.63 (15.61)	36.64 (11.18)	55.12 (13.93)	343 – 325
Letter Fluency Test	37.00 (15.36)	29.16 (12.63)	45.27 (13.55)	343–325
Token Test	33.52 (2.27)	32.61 (2.54)	34.46 (1.44)	330–318
Digit Cancellation Test	50.11 (8.94)	45.72 (9.67)	54.77 (4.82)	342–322
Digit Span Test - Forward	6.31 (1.94)	5.85 (1.22)	6.81 (1.29)	343–322
Digit Span Test - Backward	4.72 (1.34)	4.14 (1.08)	5.34 (1.31)	343–322
Visuospatial Praxis Test	12.54 (1.90)	11.80 (2.10)	13.33 (1.27)	339–321
Stroop Test - Error Interference	1.43 (4.06)	2.53 (5.20)	0.26 (1.63)	339 – 319
Stroop Test - Time Interference	26.34 (27.30)	35.07 (34.55)	17.06 (10.18)	339 – 319
Paired Associated Learning Test	13.53 (5.11)	10.93 (4.45)	16.28 (4.24	342–323
Rey-Osterrieth Figure - Copy	31.45 (5.15)	29.35 (6.01)	33.72 (2.50)	343 – 317
Rey-Osterrieth Figure - Recall	13.72 (6.82)	10.55 (5.77)	17.17 (6.19)	343–316
Logical Memory - Immediate Recall	11.28 (4.91)	9.59 (4.55)	14.92 (3.46)	336 – 156
Logical Memory - Delayed Recall	13.48 (5.87)	11.37 (5.25)	18.01 (3.61)	335–156
Raven's Progressive Matrices	30.21 (5.05)	27.75 (5.35)	32.83 (3.00)	342 – 321
Prediction 1 - Related Variables				
SIM - Quantitative Score	20.95 (5.55)	18.72 (5.13)	23.29 (4.99)	343–326
SIM - Item-Level Score - LCH Ratio	2.07 (0.11)	2.10 (0.11)	2.03 (0.10)	343–326

Means and standard deviations are indicated. Group comparisons are not reported since they were beyond the scope of the study. The two sub-cohorts, in fact, were not analysed separately but as part of a single cohort.

adjectives, adverbs, nouns and verbs as a network of interconnected nodes. A connection (named "edge") between two nodes corresponds to a lexical-semantic relation existing between two lexicon entries. Each WordNet entry is identified as a "synset", with single lexical units having multiple synsets, each expressing a distinct concept. An example of lexical unit with multiple synsets is illustrated in Box 2.

A synset was identified for each of the 38 SIM words (Table 2), in line with the concepts expressed by the correct answers (Wechsler, 2008), and each of the 19 items was assessed as a function of the semantic distance between the two synsets. This was carried out using the open-source *WordNet Similarity for Java* interface available at https://ws4jdemo.appspot.com. As a range of metrics have been proposed to quantify semantic distances based on distinct properties of synsets and their links (Supplementary Table S1), a data-driven process of selection was carried out to identify the most suitable metric (see Supplementary Material). The LCH metric quantifies the distance between two synsets via the negative logarithm of the ratio between the inter-synset shortest path length ("SPL" in the equation) and the maximum path depth (i.e., "max D" in the equation), i.e., an index of semantic specificity within the WordNet hierarchy (Fig. 2; (Leacock and Chorodow, 1998)).

$$LCH_{s1,s2} = -\log_e SPL_{s1,s2}/2 \times \max D_{SPL_{s1,s2}}$$

LCH was the best-performing operationalisation of semantic distance, as 1) none of the 19 SIM LCH metrics was equal to 0 (Supplementary Table S2); 2) LCH metrics were significantly associated with item difficulty, estimated via the proportion of correct responses across the cohort (Supplementary Table S3); and 3) the variance inflation factor (*VIF*) coefficient assessing collinearity between average LCH metrics of correct responses (calculated as a ratio between average LCH of correct responses, and number of correct responses) and standard quantitative SIM scores was only 1.601 (Supplementary Table S4), indicating a mere 37.6 % shared variance between quantitative and item-level scores (Thompson et al., 2017), confirming statistical independence of the two measures.

$$LCH\ Ratio = \sum_{Correct\ Responses} LCH \Big/ n_{Correct\ Reponses}$$

Importantly, this was in relation to fully-correct responses only (i.e., a score of 1 for SIM items 1–5, and a score of 2 for SIM items 6–19). The *VIF* coefficient was > 2 and the shared variance was ~55.3 % when LCH metrics were factored in correct and partial responses (Supplementary Table S3). For this reason, all subsequent procedures were carried out in relation to correct responses only. The association between standard SIM scores and LCH Ratio scores are shown in Fig. 3.

LCH ratio is conceptually similar to the qualitative scores described by De Marco and colleagues (De Marco et al., 2023a) in relation to Boston Naming test performance, and to those described by Vonk and colleagues (Vonk et al., 2019) and by Forbes-McKay and colleagues (Forbes-McKay et al., 2005) in relation to Category Fluency test performance: all these item-level scores consist of an average of the semantic properties associated with the items correctly named or retrieved as part of each test.

2.4. Prediction 1– Procedure and data analysis

Hierarchical regression models were designed to test whether item-level SIM scores would independently predict semantic memory performance. Three measures sensitive to semantic processing and semantic memory were selected as dependent variables from the tests listed in Table 1: Confrontational Naming, the Pyramids and Palm Trees test, and Category Fluency.

The Confrontational Naming task included 20 images extracted from the Snodgrass and Vanderwart standardised pictures (Snodgrass and Vanderwart, 1980). Participants were given a single attempt to name each image (i.e., differently from the Boston Naming test, in which participants are provided with a series of cues if the initial attempt is unsuccessful). The score on this test ranged between 0 and 20, with naming errors or omissions in this specific cohort depending on semantic and/or lexical difficulties (De Marco et al., 2023a).

The Pyramids and Palm Trees test was chosen as a task of semantic retrieval that is influenced by picture recognition (Klein and Buchanan, 2009) and is only minimally reliant on lexical or attentional/executive demands. A "three-picture" version inclusive of 52 items was administered, and the count of correct responses was extracted as a measure of semantic recognition.

Finally, Category Fluency was selected as a test of free semantic recall (Gruenewald and Lockhead, 1980). Three different categories (*cities*, *animals* and *fruits*, one minute each) were explored by each participant, and the count of correct entries (free of intrusions and perseverations) was retained as the test score. Although this test measures semantic abilities, performance is also significantly influenced by attentional/executive functioning skills (Aita et al., 2019; Whiteside

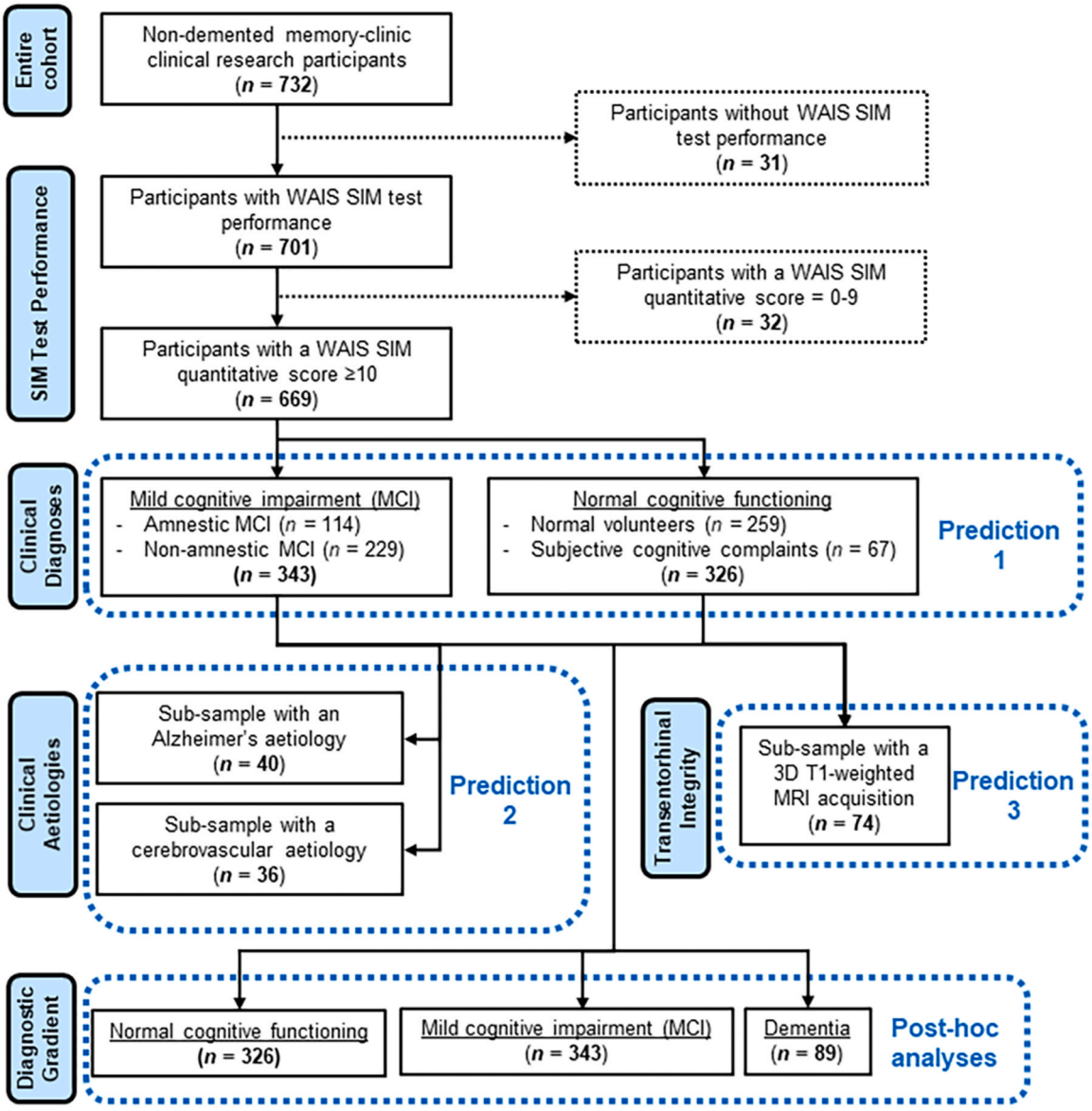


Fig. 1. Study flowchart. The flowchart illustrates the size of the cohort and sub-samples in association with each sub-study.

Box 2
An example of lexical unit with multiple synsets.

Lexical Unit: "*SKIP*"	
Synset	**Definition**
n#1	*a gait in which steps and hops alternate*
n#2	*a mistake resulting from neglect*
v#1	*bypass: 'He skipped a row in the text so the sentence was incomprehensible'*
v#2	*intentionally fail to attend: 'cut class'*
v#3	*jump lightly*
v#4	*leave suddenly; 'She persuaded him to decamp'; 'skip town'*
v#5	*bound off one point after another*
v#6	*cause to skip over a surface; 'Skip a stone across the pond'*

Table 2
SIM items and synsets selected as part of this study.

Word 1		Synset and Definition	Word 2		Synset and Definition
FORK	*n#1*	*"Cutlery used for serving and eating food"*	SPOON	*n#1*	*"A piece of cutlery with a shallow bowl-shaped container and a handle: used to stir or serve or take up food"*
SOCK	*n#1*	*"Hoisiery consisting of a cloth covering for the foot, worn inside the shoe; reaches to between the ankle and the knee"*	SHOE	*n#1*	*"Footwear shaped to fit the foot (below the ankle) with a flexible upper of leather or plastic and a sole and heel of heavier material"*
YELLOW	*n#1*	*"Yellow colour or pigment, the chromatic colour resembling the hue of sunflowers or ripe lemons"*	GREEN	*n#1*	*"Green colour or pigment, resembling the colour of growing grass"*
DOG	*n#1*	*"A member of the genus Canis (probably descended from the common wolf) that has been domesticated by that has been domesticated by man since prehistoric times"*	LION	*n#1*	*"Large gregarious predatory feline of Africa and India having a lawny coat with a shaggy mane in the male"*
COAT	*n#1*	*"An outer garment that has sleeves and covers the body from shoulder down"*	SUIT	*n#1*	*"A set of garments (usually including a jacket and trousers and skirt) for outerwear all of the same fabric and colour"*
PIANO	*n#1*	*"A keyboard instrument that is played by depressing keys that cause hammers to strike tuned strings and produce sounds"*	DRUM	*n#1*	*"A musical percussion instrument, usually consists of a hollow cylinder with a membrane stretched across each end"*
ORANGE	*n#1*	*"Round yellow to orange fruit of any of several citrus trees"*	BANANA	*n#2*	*"Elongated crescent-shaped yellow fruit with soft sweet flesh"*
EYE	*n#1*	*"The organ of sight"*	EAR	*n#1*	*"The sense organ for hearing and equilibrium"*
BOAT	*n#1*	*"A small vessel for travel on water"*	CAR	*n#1*	*"A motor vehicle with four wheels, usually propelled with an internal combustion engine"*
TABLE	*n#2*	*"A piece of furniture having a smooth flat top that is usually supported by one or more vertical legs"*	CHAIR	*n#1*	*"A seat for one person, with a support for the back"*
WORK	*n#1*	*"Activity directed toward making or doing something"*	PLAY	*n#8*	*"Activity by children that is guided more by imagination than by fixed rules"*
STEAM	*n#1*	*"Water at boiling temperature diffused in the atmosphere"*	FOG	*n#1*	*"Droplets of water vapour suspended in the air near the ground"*
EGG	*n#1*	*"Animal reproductive body consisting of an ovum or embryo together with nutritive and protective envelopes"*	SEED	*n#2*	*"A mature fertilised plant ovule consisting of an embryo and its food source and having a protective coat or testa"*
DEMOCRACY	*n#2*	*"A political system in which the supreme power lies in a body of citizens who can elect people to represent them"*	MONARCHY	*n#1*	*"An autocracy governed by a monarch who usually inherits the authority"*
POEM	*n#1*	*"A composition written in metrical feet forming rhythmical lines"*	STATUE	*n#1*	*"A sculpture representing a human or animal"*
PRAISE	*n#1*	*"An expression of approval and commendation"*	PUNISHMENT	*n#1*	*"The act of punishing"*
FLY	*n#1*	*"Two-winged insect characterised by active flight"*	TREE	*n#1*	*"A tall perennial woody plant having a main trunk and branches forming a distinct elevated crown"*
HIBERNATION	*n#1*	*"The torpid or resting state in which some*	MIGRATION	*n#4*	*"The periodic passage of*

(*continued on next page*)

Table 2 (*continued*)

Word 1		Synset and Definition	Word 2		Synset and Definition
		animals pass the Winter"			*groups of animals (especially birds or fishes) from one region to another at certain times of the year"*
ENEMY	*n#4*	*"A personal enemy"*	FRIEND	*n#1*	*"A person you know well and regard with affection and trust"*

et al., 2016).

Three additional tasks of executive functioning were also selected from the battery of tests listed in Table 1. These were the Digit Span test (backwards presentation), the Stroop test (time interference) and the Letter Fluency test. To test independence of LCH metrics at predicting performance on the three semantic tests from executive functioning and from standard SIM scores, three hierarchical regression blocks were set up. Demographic and clinical characteristics (i.e., diagnosis, age, years of education, sex, Mini Mental State Examination score, and scores on the three executive tasks) were included in Block 1, standard SIM scores were included in Block 2, and LCH measures were included in Block 3. To test whether Block 3 would predict a statistically significant portion of outcome variability, each block was evaluated as a function of its r^2 coefficient, and Block 2-to-Block 3 r^2-*change* was tested with an *F*-test for nested model comparisons. Three regression models were defined (one per semantic test), and the threshold of Block 2-to-Block 3 significance was set to $p < 0.05$. A graphic representation of these analyses is illustrated in Fig. 4. The IBM SPSS software platform, version 29.0.1.0 (https://www.ibm.com/spss) was used to run these analyses.

2.5. Prediction 2– Procedure

The objective of this analysis was to test the statistical impact of LCH Ratio on clinical aetiologies. In fact, AD induces a progressive degeneration of semantic processing (Venneri et al., 2016, 2018) and, while

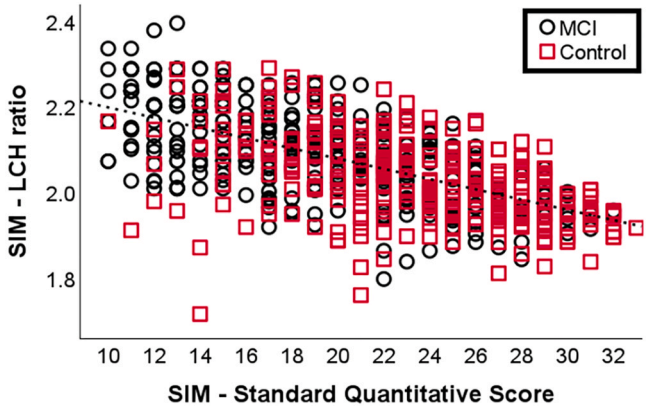


Fig. 3. Linear association between standard quantitative SIM scores and the index of semantic distance used in this study: LCH Ratio. The equation corresponding to the dotted line equals to $LCH\ Ratio = 2.319 + (-0.012) \times quantitative\ SIM\ score$.

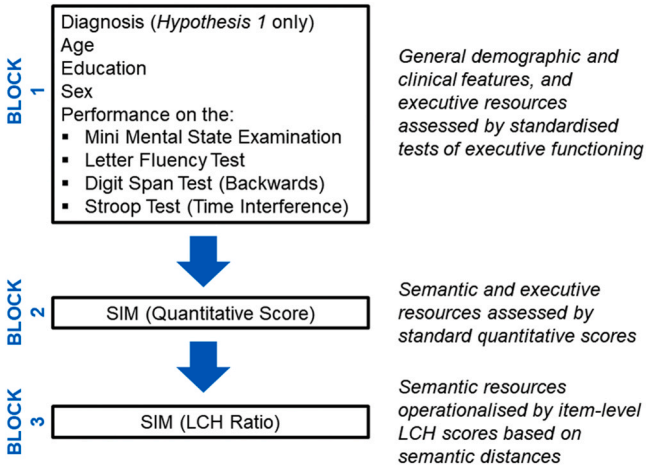


Fig. 4. Visual and conceptual representation of hierarchical regression models designed to test *Prediction 1* and *2*. "Diagnosis" was not included in the model testing *Prediction 2*. The same variables (plus total intracranial volumes and grey matter ratios) were also used as part of the approach addressing *Prediction 3*.

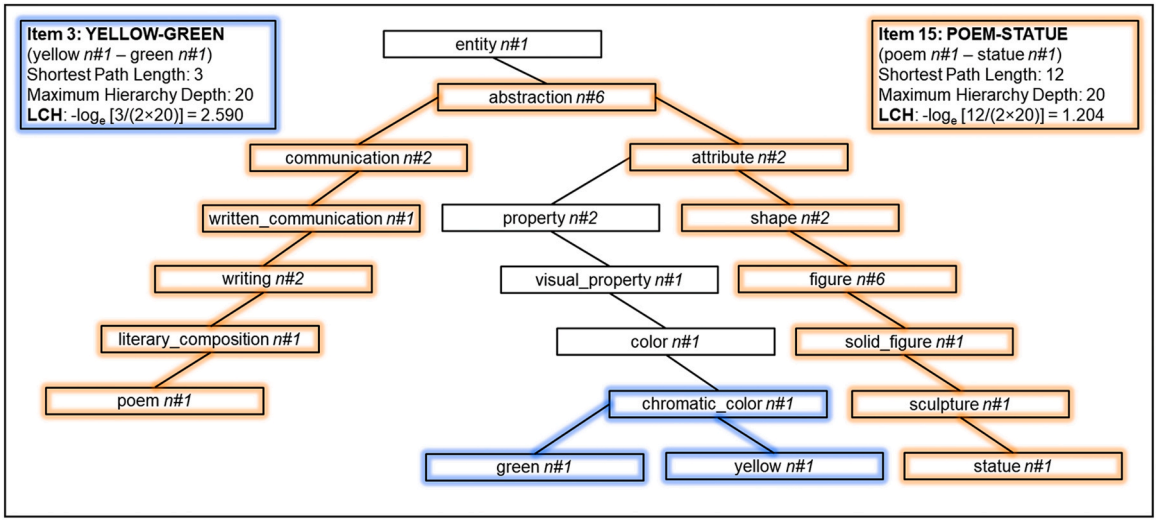


Fig. 2. Visual representation of the LCH metric. The distance path for two SIM items is illustrated. A first path characterised by a short distance (i.e., YELLOW-GREEN) is shown in blue and corresponds to an LCH score of 2.590. A second path characterised by a longer distance (i.e., POEM-STATUE) is shown in orange and corresponds to a LCH score of 1.204.

quantitative SIM scores might not be sufficiently sensitive to this decline (due to them being influenced by executive functioning), item-level scores factoring in semantic distances might be more sensitive to subtle semantic decline. Two sub-samples of MCI individuals with a clinically established aetiology were identified from the original cohort: 40 with AD and 36 with cerebrovascular disease. Aetiologies were established by a consensus of clinicians, followed the recommendations of UK's National Institute for Health and Care Excellence, and were confirmed by follow-up assessments indicating worsening of clinical symptoms. Moreover, none of the participants with a cerebrovascular aetiology had suffered major brain infarcts. The demographic and clinical profile of these two clinical sub-samples is illustrated in Table 3. A hierarchical logistic regression was run to predict aetiological statuses via a model comparable to those designed to address *Prediction 1* (i.e., Block 1: age, years of education, sex, Mini Mental State Examination score, and scores on the three executive tasks; Block 2: Quantitative SIM scores; Block 3: LCH ratio). As with *Prediction 1* testing, executive tasks were included as control factors to evaluate the statistical impact of LCH ratio net of the effect of executive predictors (Fig. 4). As all cases included in this analysis were MCI individuals, diagnosis was not added as a predictor in this model. The IBM SPSS software platform, version 29.0.1.0 (https://www.ibm.com/spss) was used to run these analyses.

2.6. Prediction 3– Procedure

The possibility of LCH metrics being a predictor of anatomical integrity of the sub-hippocampal cortex was tested in a subgroup of 74 participants with anatomical MRI of the brain (i.e., 42 cognitively healthy adults and 32 participants with MCI). Three-dimensional T1-weighted images were acquired on a 3 Tesla Philips Ingenia MRI scanner with the following technical specifications: 0.94 × 0.94 × 1.00 mm voxel dimension, 8.2 s repetition time, 3.8 s echo delay time, 256 mm field of view, and 256 × 256 × 170 matrix size.

Table 3
Characterisation of sub-sample addressing study *Prediction 2*.

	AD	Vascular	Significance
Demographic/Clinical Variable			
Age (years)	66.30 (10.71)	66.33 (9.37)	n.s.
Education (years)	12.90 (2.86)	12.36 (3.45)	n.s.
Sex (F/M)	25/15	15/21	n.s.
Mini Mental State Examination	26.23 (1.47)	26.50 (1.56)	n.s.
Confrontational Naming	19.28 (0.75)	19.25 (0.73)	n.s.
Pyramids & Palm Trees Test	50.10 (1.31)	50.47 (1.32)	n.s.
Category Fluency Test	34.80 (10.42)	33.72 (8.29)	n.s.
Letter Fluency Test	32.45 (11.51)	23.28 (12.11)	$p = 0.001$
Token Test*	32.66 (2.37)	31.97 (2.78)	n.s.
Digit Cancellation Test	47.50 (8.09)	43.39 (9.29)	$p = 0.043$
Digit Span Test - Forward	5.80 (1.09)	5.39 (0.80)	n.s.
Digit Span Test - Backward	4.30 (1.11)	3.86 (0.90)	n.s.
Visuospatial Praxis Test	11.75 (1.76)	11.86 (2.21)	n.s.
Stroop Test - Error Interference	3.67 (6.15)	3.21 (6.40)	n.s.
Stroop Test - Time Interference	34.01 (20.49)	35.19 (16.55)	n.s.
Paired Associated Learning Test	9.65 (4.39)	10.17 (4.50)	n.s.
Rey-Osterrieth Figure - Copy	29.92 (5.10)	28.94 (6.48)	n.s.
Rey-Osterrieth Figure - Recall	7.75 (4.90)	9.79 (5.04)	n.s.
Logical Memory - Immediate Recall	6.73 (4.39)	9.94 (4.12)	$p = 0.002$
Logical Memory - Delayed Recall	6.98 (5.12)	11.67 (5.00)	$p < 0.001$
Raven's Progressive Matrices	28.26 (4.42)	27.00 (4.62)	n.s.
Prediction 2-Related Variables			
SIM - Quantitative Score	18.78 (5.49)	18.39 (4.65)	n.s.
SIM - Item-Level Score - LCH Ratio	2.13 (0.10)	2.08 (0.11)	$p = 0.027$

Means and standard deviations are indicated. *two participants with a diagnosis of AD did not have any score available for this test. One-sided *p*-values are reported in relation to *Prediction 2*-related variables, to reflect the directionality of the study hypothesis. Two-sided *p*-values are reported in relation to the remaining variables.

To analyse grey matter density, images were processed with voxel-based morphometry procedures via Statistical Parametric Mapping 12 (Wellcome Centre for Human Neuroimaging, London, UK), running in a MATLAB environment (MathWorks, version R2021a). Probabilistic segmentation (inclusive of spatial modulation and normalisation) was run to separate and extract three distinct tissue-class sub-maps (i.e., grey matter, white matter and cerebrospinal fluid). These three sub-maps were quantified in their native space (in ml) using the *get_totals* command line (http://www0.cs.ucl.ac.uk/staff/G.Ridgway/vbm/get_totals.m). The three tissue-class volumes were then summed up to obtain individual indices of total intracranial volume. The fraction of grey matter volume (i.e., grey matter volumes divided by total intracranial volumes) were also calculated. Grey matter sub-maps were finally smoothed with an 8 mm full-width at half-maximum Gaussian kernel.

Linear models were designed to test the voxel-by-voxel association between SIM scores (i.e., quantitative and item-level), and grey matter density. Following the framework by Hyatt and colleagues (Hyatt et al., 2020) on the role of covariates in voxel-based analyses of neuroanatomical MRI images, uncorrected models were initially launched. A principal component analysis was then run to optimise the number of continuous covariates, given the very large number of statistical models on which voxel-wise whole-brain analyses are based. Two components (a cognitive/reserve one, and an ageing/integrity one, see Supplementary Material) accounting for age, years of education, Mini Mental State Examination scores, total intracranial volumes (i.e., a recognised index of brain reserve), grey matter ratios (i.e., an index of global atrophy), performance on the three executive tests described in the previous sections, and quantitative SIM scores (this last one only as part of LCH analyses) were added as covariates as part of corrected models (Supplementary Fig. S1). Diagnosis and sex were added as further covariates in these models, i.e., since categorical variables are not processable by principal component analyses. A cluster-forming threshold of $p < 0.005$ was applied to the resulting statistical maps, and clusters were reported as statistically significant when surviving both Family-Wise Error and False Discovery Rate-corrected *p*-values < 0.05. Talairach coordinates were converted to the Montreal Neurological Institute (MNI) space via a non-linear transformation (https://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach), and were interpreted using the Tailarach Daemon applet (Lancaster et al., 2000).

The same volumetric 74 T1-weighted images were also processed with the FreeSurfer version 7.1.1 open-source package (Athinoula A. Martinos Center for Biomedical Imaging, Harvard, USA) to analyse the association between item-level SIM scores and region-of-interest trans-entorhinal cortical thickness. The standard recon-all procedure was followed, as previously described (Dale et al., 1999; Fischl and Dale, 2000; Fischl et al., 2002, 2004). This consists of intensity normalisation, removal of nonbrain tissue, segmentation, surface inflation, and topological correction. Cortical thickness was calculated as the closest distance from the grey/white boundary to the grey/cerebrospinal fluid boundary at each vertex on the surface. In particular, we automatically estimated separate measures for the left and right hemispheres of the pars opercularis, pars triangularis, entorhinal, and perirhinal cortex from the *ex vivo* parcellations, where the labels were defined cytoarchitecturally from a set of *post mortem* data, and then averaged and thresholded as previously described (Augustinack et al., 2013, 2014). The output was quality-checked for imprecise segmentations and, as a result, 6 scans (5 cognitively healthy adults and 1 participant with MCI) were excluded, leaving a sample of $n = 68$ for this analysis. Correlations and hierarchical regressions were defined to mirror the uncorrected and fully-corrected voxel-by-voxel models described above, respectively. The entorhinal and perirhinal cortex were the regions of interest of these analyses, while Broca area and its contralateral counterpart (i.e., pars opercularis and pars triangularis) were selected as control regions. To

correct for multiple comparisons (8 in total), a Bonferroni-corrected $p < 0.00625$ threshold of significance was applied to these analyses.

3. Results

3.1. Item-level SIM scores as predictors of semantic memory performance

These findings are shown in Table 4. Quantitative SIM scores were a significant predictor of performance on all three tests of semantic memory. LCH Ratio predicted an additional small but significant portion of variability of Confrontational Naming and Pyramids and Palm Trees test performance (1.7 % and 1.5 % of additional variance, respectively). Block 1 findings indicated that these tests were only modestly predicted by performance on the three executive tests (i.e., the average absolute magnitude of their standardised slopes was 0.057 and 0.078, in relation to Confrontation Naming and Pyramids and Palm Trees, respectively).

LCH Ratio was not a significant predictor of Category Fluency scores. These were instead significantly predicted by all three executive tests (the average of the three standardised slopes for the executive tests was 0.205, considerably larger than those of the two semantic tests described above). In summary, these findings support *Prediction 1* of the study, although not in relation to a semantic memory test that is significantly supported by executive functioning.

3.2. Item-level SIM scores and clinical aetiologies

The two groups of MCI individuals did not differ in age, education, sex, or general cognitive status. The group with an AD aetiology showed worse memory performance (Logical Memory Test), while the group with a cerebrovascular aetiology showed worse performance in tests of visuospatial attention and speed of processing (Digit Cancellation Test) and lexical-executive skills (Letter Fluency test). While the two groups did not differ in quantitative SIM performance, the group with an AD aetiology showed a higher LCH Ratio, indicating an average lower semantic distance (Table 3).

The outcome of the logistic regression model is outlined in Table 5. To ensure the validity of the model, we conducted a Box-Tidwell test for the predictor of interest: an interaction term was computed between LCH Ratio and its natural logarithmic transformation, and this was included into the final block of the model. The result showed a non-significant coefficient for the interaction term ($b = 127.83$, $p = 0.136$), indicating that the assumption of linearity in the logit was met. Compared with the null model, Block 1 significantly improved the prediction of aetiological status ($\chi^2(7) = 15.70$, $p = 0.028$), with performance on the Letter Fluency test being the only significant predictor ($b = -0.072$, $p = 0.006$). Although Block 2 (i.e., quantitative SIM scores) did not offer any significant contribution ($\chi^2(1) = 1.88$, $p = 0.170$), Block 3 marked an improvement, with the final predictor, LCH ratio, significantly increasing the fit of the model ($\chi^2(1) = 7.22$, $p = 0.007$). None of the Hosmer and Lemeshow statistics was significant, indicating that the model was well calibrated. Table 5 includes all Nagelkerke Pseudo r^2 coefficients and odds-ratio values. These findings are in support of *Prediction 2*. As the two sub-cohorts showed statistically significant differences in performance on three additional neuropsychological scores (i.e., Digit Cancellation Test, Logical Memory Test – Immediate Recall, and Logical Memory Test – Delayed Recall; Table 3), the logistic regression model was also run including these three variables as part of Block 1. With these three additional variables, the prediction offered by Block 1 improved considerably ($\chi^2(10) = 31.67$, $p < 0.001$), but the outcome of Block 2 ($\chi^2(1) = 0.68$, $p = 0.409$) and Block 3 ($\chi^2(1) = 6.69$, $p = 0.010$) did not change.

Table 5
Outcome of the logistic regression model designed to test *Prediction 2*.

	Block 1		Block 2		Block 3	
χ^2	15.697[a]		1.88		7.222*	
Nagelkerke pseudo r^2	0.249		0.276		0.372	
	Log Odds	Odds Ratio	Log Odds	Odds Ratio	Log Odds	Odds Ratio
Age	0.014	1.014				
Education	0.039	1.040				
Sex	−0.655	0.519				
MMSE	0.224	1.250				
Letter Fluency	−0.072*	0.930				
Digit Span Test - Backwards	−0.250	0.779				
Stroop Time Interference	−0.012	0.988				
SIM Score			0.088	1.092		
SIM LCH Ratio					−8.906[a]	0.000136

Odds indicate that for each increase in test scores, there is an indicated proportional decrease in the odds of having AD as underlying aetiology. The Odds Ratio for Letter Fluency had a 95 % confidence interval of [0.884, 0.979] and the score indicated ~7 % decreased odds of having AD per each + 1 increase in test performance. The Odds Ratio for LCH Ratio had a 95 % confidence interval of [0.00000014, 0.134] and the coefficients indicated a ~1.36 % increased odds of having AD per each + 0.01 increase in this item-level score. As sex was not a significant predictor, the reference level is not indicated.

* $p < 0.01$.

[a] $p < 0.05$;

Table 4
Hierarchical regression models testing the independent association between item-level SIM scores (i.e., LCH ratio) and performance on established semantic memory tests. Standardised slope coefficients are indicated in association with each predictor.

Outcome Test	Block	r^2[a]	Diagnosis	Age	Education	Sex	MMSE	Letter Fluency	Digit Span Backwards	Stroop Time Interference	SIM Score	SIM LCH Ratio
Naming	1	0.070***	0.001	0.140***	−0.112[b]	−0.056	0.188***	0.099[b]	0.045	−0.028		
	2	0.085**									0.162**	
	3	0.102***										−0.167***
Pyramids & Palm Trees	1	0.086***	0.058	0.204***	0.010	−0.015	0.104[b]	0.112[b]	−0.009	−0.113*		
	2	0.126***									0.260***	
	3	0.141**										−0.155**
Category Fluency	1	0.577***	0.228***	0.022	0.092**	−0.026	0.123***	0.426***	0.066[b]	−0.069[b]		
	2	0.586***									0.129***	
	3	0.587										−0.025

* $p < 0.01$;

** $p < 0.005$;

*** $p < 0.001$;

[a] *p*-values of the *r-square* change are indicated. As sex was not a significant predictor, the reference level is not indicated.

[b] $p < 0.05$;

3.3. Item-level SIM scores as predictors of structural integrity of the sub-hippocampal cortex

No significant association was found between quantitative SIM scores and grey matter in the uncorrected analysis. This was thus not followed up with a corrected model.

The uncorrected model indicated a negative association between LCH ratio and a large brain area including the frontal, temporal and limbic territory (Supplementary Fig. S2). No positive associations were found in this analysis. Corrected analyses confirmed the presence of a negative statistical association (i.e., the lower the score, the higher density) in the parahippocampal gyrus, with peaks of significance in the perirhinal cortex, bilaterally. Additional peaks (i.e., z-scores > 3.5) were found between LCH ratio and grey matter density in the superior temporal gyrus, bilaterally, in the right inferior frontal lobe, in the cerebellar culmen, and in a right cerebellar cluster centred in Lobule VI (Table 6; Fig. 5). These clusters also extended to the left entorhinal cortex and fusiform gyrus, right superior temporal gyrus, precentral gyrus (bilaterally), and insula (bilaterally). No positive associations were found.

None of the correlational and hierarchical regression models indicated an association between regional cortical thickness and quantitative or item-level qualitative SIM scores.

These findings collectively support *Prediction 3* of the study, although only in relation to perirhinal density (i.e., not thickness). Moreover, the analyses revealed the involvement of other regions external to the mediotemporal lobe.

3.4. LCH ratio in participants with dementia

A group of 89 participants with a diagnosis of dementia, a Mini Mental State Examination score between 21 and 23 (indicating a very mild level of severity), and a quantitative SIM score > 9 was identified among the outpatients recruited at the same institution and within the same timeframe as the main cohort (Fig. 1). Their LCH ratio was calculated and a one-way *ANOVA* was run to compare quantitative and item-level SIM scores across diagnostic statuses. Levene's test statistic indicates homogeneity of variance across the three groups for both measures. A significant effect of group was found for both quantitative ($F_{2,\ 757} = 105.945$, $p < 0.001$) and item-level ($F_{2,\ 757} = 45.106$, $p < 0.001$) scores. Quantitative SIM scores showed a progressive decline in quantitative performance, with Bonferroni-corrected *post-hoc* tests indicating a statistically significant difference across all three groups (Fig. 6**A**). *Post-hoc* analyses indicated that item-level SIM scores showed a difference between healthy controls and each clinical group, but no difference between participants with MCI and participants with dementia (Fig. 6**B**). As two observations fell below 1.5 of their interquartile range and were thus flagged up as potential outliers in the distribution of LCH Ratio scores for healthy controls (i.e., a score of 1.718) and for participants with dementia (i.e., as score of 1.843), the analyses were rerun after removing these two data points. The pattern of findings was replicated (at the same statistical thresholds) following this adjustment.

4. Discussion

Described since the end of the nineteenth century (Binet and Henri, 1895), SIM has been widely used, in isolation or as part of the extended WAIS battery, as a measure of intelligence and cognitive proficiency. The current consensus, outlined by the WAIS technical manual (Wechsler, 2008) and confirmed by data-driven factor analyses (Merz et al., 2021), indicates that SIM performance loads on a "Verbal Comprehension" factor, together with Vocabulary and Information, two additional sub-tests of crystallised intelligence based on knowledge retrieval. Cohort-based evidence indicates that SIM performance starts declining during the earliest stages of AD (Amieva et al., 2005; Elias et al., 2000; Jacobs et al., 1995), in line with the view of semantic memory decline being a major clinical phenotype at the AD preclinical stage (Venneri et al., 2016, 2018). While the traditional quantitative scoring approach only partially accounts for item difficulty, an item-level procedure based on path length can capture an individual's ability to navigate through the lexical-semantic network with more precision and with minimal influence of executive functioning. In this study, this item-level operationalisation of semantic memory processing was independently associated with performance on two consolidated tests of semantic memory (Confrontational Naming and Pyramids and Palm Trees tests), beyond the portion of variability accounted for by classic SIM scores. Conversely, the LCH ratio was not a significant predictor of performance on a test of semantic memory that is significantly influenced by several executive-functioning processes (Category Fluency test). Access to semantic knowledge is mediated by automatic and controlled processes, and different semantic tests require different amounts of control resources (Arroyo-Anlló et al., 2011). Item-level scores of semantic memory tests tend to be less influenced by controlled processes (De Marco et al., 2023b, 2025), and, in this respect, the lack of association with the Category Fluency test scores confirms this principle.

The findings described in **Section 3.3** indicate that the LCH ratio was significantly negatively associated with grey matter density in the sub-hippocampal portion of the mediotemporal lobe, bilaterally. This is the earliest cortical territory affected by hyperphosphorylated tau pathology in AD (Braak and Braak, 1995; Braak et al., 2006; Igarashi, 2023). These neuropathological changes are associated with local neurodegenerative changes in this region (Krumm et al., 2016; Xie et al., 2018), and for this reason, atrophy of this region can be considered an indirect marker of tau pathology (Dallaire-Théroux et al., 2019). In this context, an item-level scoring approach to SIM can provide

Table 6
Whole-brain analysis of grey matter maps (corrected model).

Cluster Number	Cluster-Level p_{FWE}-value	Cluster-Level p_{FDR}-value	Cluster Extent (voxels)	Z-Score at Local Maximum	Brodmann Area	Hemisphere	Brain Region	Tailarach Coordinates		
								x	y	z
LCH Ratio-Grey-Matter Density - Negative Association ($p_{CFT} = 0.005$)										
1	0.009	0.013	2032	3.93	38	L	Superior Temporal Gyrus	−45	2	−8
				3.55	42	L	Transverse Temporal Gyrus	−56	−14	9
2	0.012	0.013	1898	3.84	35	R	Parahippocampal Gyrus	22	−22	−19
3	0.025	0.020	1636	3.72		L	Cerebellum - Culmen	−20	−32	−15
				3.70	35	L	Parahippocampal Gyrus	−21	−22	−17
				3.54	11	L	Cerebellum - Culmen	−20	−44	−11
4	0.003	0.008	2493	3.69	47	R	Inferior Frontal Gyrus	27	14	−18

CFT: Cluster-Forming Threshold; FDR: False Discovery Rate; FWE: Family-Wise Error; L: Left; R: Right

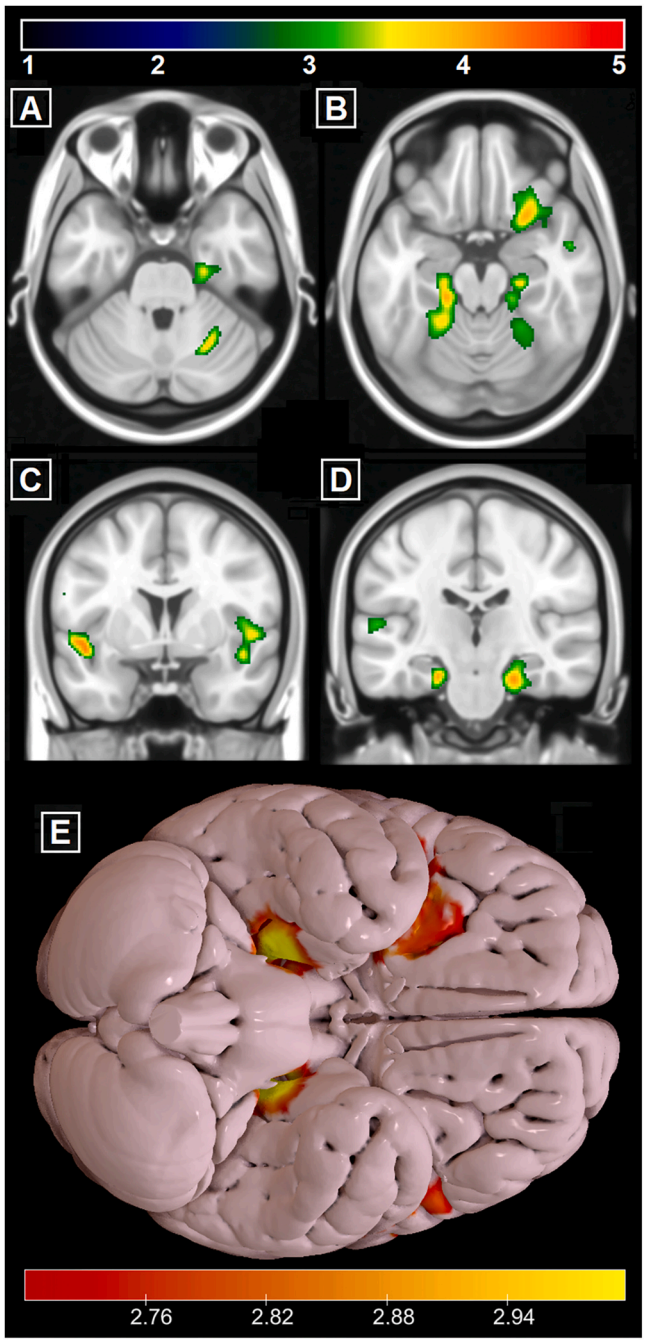


Fig. 5. Statistical negative association (significant at a cluster-forming threshold of $p < 0.005$, and at a cluster-level Family-Wise Error and False Discovery Rate-corrected $p < 0.05$) between grey matter density and average LCH ratio. *z*-scores are colour-coded according to the scales reported at the top and at the bottom. MNI slices are as follows, 'A': y = -22; 'B': y = 2; 'C': z = -20; 'D': z = -30. The MNI-152 T1 (0.5 mm) template was used to visualise the statistical map, and the MRIcron image viewer was used to visualise the template and the overlay. Perirhinal peaks (i.e., Brodmann area 35) are shown in 'D', while 'B' shows these mediotemporal clusters extending to the superior portion of the cerebellum (i.e., culmen). The bilateral involvement of the portion of the superior temporal gyrus located in proximity of the posterior planum temporale is shown in 'C' (the right cluster also extends to the right inferior frontal lobe, as shown in 'B'), while 'A' shows the cerebellar cluster located in correspondence to Lobule VI. All slices are in the neurological visualisation. 'E', finally, is a view of the basal portion of the contrast superimposed on the "*mni152_2009.curv*" template that highlights the involvement of the parahippocampal gyrus, bilaterally. The Surf Ice surface-rendering tool (https://www.nitrc.org/projects/surfice/) was used to this aim.

neuropsychologists with a clinical tool that is susceptible to pathological and neurodegenerative changes affecting the entorhinal and perirhinal cortex. Processing of semantic relationships between two concepts is supported by activity of the perirhinal cortex (Bruffaerts et al., 2019; Lambon Ralph et al., 2017). As longer WordNet distances define more complex semantic relationships (Supplementary Table S3), perirhinal atrophy would result into impoverished connectivity with cortical regions where concepts representations are stored (i.e., see (Huth et al., 2016), for evidence of concepts representations stored throughout the entire cortex) and, as a result, would hinder access to semantic information for the purpose of the SIM task. As perirhinal atrophy is observed in early-stage AD, this is also in line with the findings reported in **Section 3.2**: while no difference exists in quantitative SIM scores between individuals with MCI due to AD and individuals with MCI of a cerebrovascular origin, the LCH ratio was significantly higher (indicating poorer semantic performance) in the AD group, and was a significant predictor of aetiological status at the MCI stage.

The voxel-based analyses revealed a pattern of additional temporal, frontal and cerebellar regions associated with the LCH ratio. Item-level scores were bilaterally associated with the planum temporale. While the left planum temporale plays a role in language comprehension (as it partially overlaps with Wernicke area, i.e., (Shapleske et al., 1999)) and is part of the language neuroarchitecture (Hickok, 2009), the contralateral territory is involved in auditory attention and stimulus selection (Hirnstein et al., 2013). A significant negative association was also found in the right inferior frontal cortex (i.e., Brodmann Area 47). This region plays a key role in creating and updating mental representations that depend on semantic information (Tesink et al., 2009a) and in unifying this semantic information (Menenti et al., 2009; Tesink et al., 2009b). A further significant cluster was found in the right cerebellum, in correspondence with Lobule VI (Supplementary Fig. S3). This is the cerebellar region with the most significant involvement in expressive language and verbal working memory, as reported by activation likelihood estimation meta-analytical evidence (Keren-Happuch et al., 2014). Although a more recent meta-analysis indicates that this cerebellar lobule is more involved in phonological, rather than semantic processing (Turker et al., 2023), other evidence highlights its role in the processing of semantic distance between words (Lundin et al., 2023). It is particularly interesting to remark that the association between semantic cognition and cerebral resources has emerged from multimodal imaging evidence, including clinical neuro-anatomical MRI studies, magnetoencephalographic recordings, Positron Emission Tomography imaging, neuromodulation techniques, and task-based functional MRI data collected on healthy research participants. The peak regions reported in Table 6 also extended to a wider territory including the left entorhinal region, left fusiform gyrus, right superior temporal gyrus, bilateral insula and bilateral precentral gyrus, all regions that support aspects of processing that can support SIM performance, such as semantic control (Jackson, 2021), motoric imagery (Hétu et al., 2013), knowledge of living things (Thompson-Schill et al., 1999), or learning of a semantic rule (Liu et al., 2023). Taken together, this evidence indicates that qualitative SIM scores are associated with regions that sustain cognitive abilities linked to linguistic processing, some of which are known to be affected by early-AD pathology. On this note, it is also interesting to note that standard quantitative SIM scores were not associated with grey matter, in spite of the relatively liberal cluster-forming significance threshold selected for the analyses. In all likelihood, this is due to the heterogeneous set of skills testees can rely on to engage in the task, and the diverse range of cognitive operations that can lead to formulating an answer. In summary, the combination of results is in support of the general study hypothesis, as LCH ratio demonstrated to predict central neurocognitive aspects of clinical relevance associated with semantic memory processing, after controlling for standard quantitative SIM scores.

This study is not free from limitations. A first, methodological point is the fact that the correct response to each SIM trial is defined by the

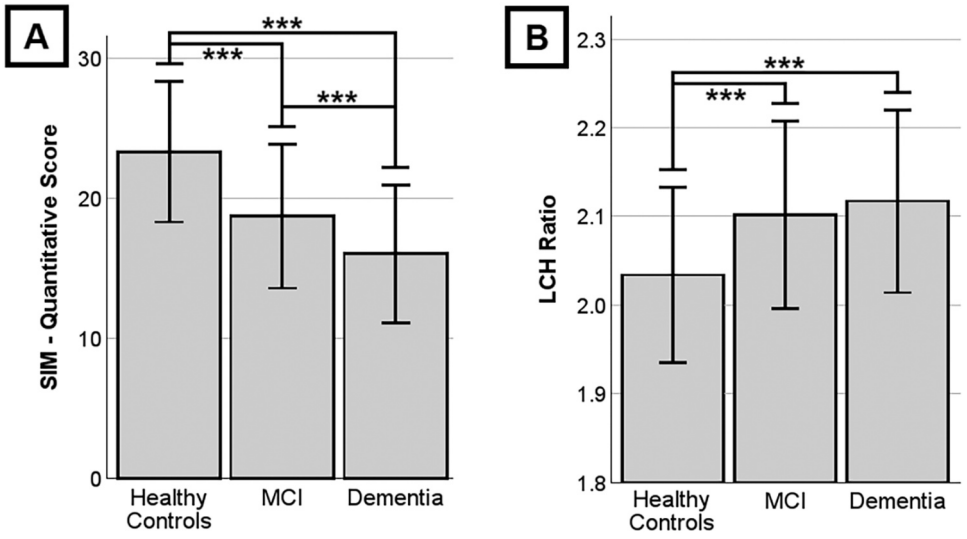


Fig. 6. Clinical gradient of quantitative ('A') and item-level ('B') SIM scores across diagnostic statuses. ***: $p < 0.001$. A one-way *ANOVA* and Bonferroni-corrected *post-hoc t*-test comparisons were used to compare the three diagnostic groups. Error bars indicate 1 SD dispersion.

WAIS scoring rules. WordNet paths and each path's *Lowest Common Subsumer* (i.e., the "ancestor" both synsets have in common, see Supplementary Table S1), however, do not always correspond to the correct response. While this is important from a construct-validity viewpoint, however, it does not influence the external validity of the LCH Ratio as a measure of semantic difficulty. In fact, a significant correlation (i.e., ~ 0.5) existed between trial LCH and the proportion of correct answers, indicating that the metric used in this study was sensitive to item difficulty (Supplementary Table S1). We propose that other methodologies, not necessarily based on a linguistic hierarchy (such as those based on Natural Language Processing and vector similarity) might be valid alternatives to the operationalisation of semantic distance. A second, conceptual aspect is the exclusive reliance on clinical diagnoses without the support offered by biomarkers. This limitation would almost exclusively affect the analyses testing *Prediction 2*, as an aetiology of AD (or cerebrovascular disease) was used as an outcome in this analysis, i.e., diagnostic status was only used as a correction factor in the analyses addressing *Prediction 1* and *3*. Thirdly, it is important to recognise that this study was carried out in relation to WAIS-III SIM items, as this allowed us to rely on a large, "historical" clinical cohort. Although subsequent WAIS editions have introduced novel SIM items, this does not invalidate the use of previous versions of the task, and of the procedures at the foundation of this study. Finally, although index definition and hypothesis testing were carried out in the same cohort, *post-hoc* analyses included an independent sub-cohort, ruling out, at least in part, biases due to potential overfitting.

In conclusion, the evidence presented in this study indicates that an item-level score of SIM performance based on the semantic distance between the two words can be very informative to cognitive neuroscientists involved in clinical research. In fact, this score is an independent predictor of semantic processing abilities, perirhinal integrity, and of AD aetiological status at the MCI stage. While standard quantitative scores are useful at characterising semantic processing and executive thinking abilities, relying on an item-level score can provide additional information (that would otherwise be ignored) that is considerably less influenced by executive functioning, and that might be of clinical relevance in the early detection of AD. Moreover, clinical applicability may also extend to other conditions in which characterisation of semantic abilities is of interest, e.g., semantic dementia.

Funding Details

This work was supported by a grant from the Alzheimer's Association (23AARG-1030190), awarded to MDM. AV is supported by funding obtained under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.3 - Call for tender No. 341 of 15/03/2022 of Italian Ministry of University and Research funded by the European Union – NextGenerationEU, Project code PE0000006, Concession Decree No. 1553 of 11/10/2022 adopted by the Italian Ministry of University and Research, CUP D93C22000930002, "A multiscale integrated approach to the study of the nervous system in health and disease" (MNESYS).

CRediT authorship contribution statement

Annalena Venneri: Writing – review & editing, Resources, Project administration, Funding acquisition, Data curation. **Satyam Chauhan:** Writing – review & editing, Data curation. **Martina Bocchetta:** Writing – review & editing, Methodology, Formal analysis. **Matteo De Marco:** Writing – original draft, Visualization, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.brainresbull.2025.111439.

Data availability

The dataset underpinning this publication can be accessed from the European Platform for Neurodegenerative Diseases (EPND), at https://discover.epnd.org/catalogue/datasets/a6430adf-d6d2-4d92-8a74-a37ca0c4cef8.

References

Aita, S.L., Beach, J.D., Taylor, S.E., Borgogna, N.C., Harrell, M.N., Hill, B.D., 2019. Executive, language, or both? An examination of the construct validity of verbal fluency measures. Appl. Neuropsychol. Adult 26, 441–451. https://doi.org/10.1080/23279095.2018.1439830.

Amieva, H., Jacqmin-Gadda, H., Orgogozo, J.M., Le Carret, N., Helmer, C., Letenneur, L., Barberger-Gateau, P., Fabrigoule, C., Dartigues, J.F., 2005. The 9 year cognitive decline before dementia of the Alzheimer type: A prospective population-based study. Brain 128, 1093–1101. https://doi.org/10.1093/brain/awh451.

Amieva, H., Le Goff, M., Millet, X., Orgogozo, J.M., Pérès, K., Barberger-Gateau, P., Jacqmin-Gadda, H., Dartigues, J.F., 2008. Prodromal Alzheimer's disease: Successive emergence of the clinical symptoms. Ann. Neurol. 64, 492–498. https://doi.org/10.1002/ana.21509.

Arroyo-Anlló, E.M., Bellouard, S., Ingrand, P., Gil, R., 2011. Effects of automatic/controlled access processes on semantic memory in Alzheimer's disease. J. Alzheimers Dis. 25, 525–533. https://doi.org/10.3233/jad-2011-110083.

Augustinack, J.C., Huber, K.E., Stevens, A.A., Roy, M., Frosch, M.P., van der Kouwe, A.J. W., Wald, L.L., Van Leemput, K., McKee, A.C., Fischl, B., for the Alzheimer's Disease Neuroimaging Initiative, 2013. Predicting the location of human perirhinal cortex, Brodmann's area 35, from MRI. Neuroimage 64, 32–42. https://doi.org/10.1016/j.neuroimage.2012.08.071.

Augustinack, J.C., Magnain, C., Reuter, M., van der Kouwe, A.J.W., Boas, D., Fischl, B., 2014. MRI parcellation of ex vivo medial temporal lobe. Neuroimage 93, 252–259. https://doi.org/10.1016/j.neuroimage.2013.05.053.

Binder, J.R., Desai, R.H., 2011. The neurobiology of semantic memory. Trends Cogn. Sci. 15, 527–536. https://doi.org/10.1016/j.tics.2011.10.001.

Binet, A., Henri, V., 1895. La psychologie individuelle, 1895 L'Année Psychol. 2, 411–465. https://doi.org/10.3406/psy.1895.1541.

Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H., Del Tredici, K., 2006. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. Acta Neuropathol. 112, 389–404. https://doi.org/10.1007/s00401-006-0127-z.

Braak, H., Braak, E., 1995. Staging of Alzheimer's disease-related neurofibrillary changes. Neurobiol. Aging 16, 271–278. https://doi.org/10.1016/0197-4580(95)00021-6.

Bruffaerts, R., De Deyne, S., Meersmans, K., Liuzzi, A.G., Storms, G., Vandenberghe, R., 2019. Redefining the resolution of semantic knowledge in the brain: Advances made by the introduction of models of semantics in neuroimaging. Neurosci. Biobehav. Rev. 103, 3–13. https://doi.org/10.1016/j.neubiorev.2019.05.015.

Clarke, A., Tyler, L.K., 2014. Object-specific semantic coding in human perirhinal cortex. J. Neurosci. 34, 4766–4775. https://doi.org/10.1523/jneurosci.2828-13.2014.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9, 179–194. https://doi.org/10.1006/nimg.1998.0395.

Dallaire-Théroux, C., Beheshti, I., Potvin, O., Dieumegarde, L., Saikali, S., Duchesne, S., for the Alzheimer's Disease Neuroimaging Initiative, 2019. Braak neurofibrillary tangle staging prediction from in vivo MRI metrics. Alzheimers Dement. D. A. D. M. 11, 599–609. https://doi.org/10.1016/j.dadm.2019.07.001.

De Marco, M., Bocchetta, M., Venneri, A., for the Alzheimer's Disease Neuroimaging Initiative, 2023a. Item-level scores on the Boston Naming Test as an independent predictor of perirhinal volume in individuals with mild cognitive impairment. Brain Sci. 13, 806. https://doi.org/10.3390/brainsci13050806.

De Marco, M., Vonk, J.M.J., Quaranta, D., 2023b. Editorial: The mechanistic and clinical principles of item-level scoring methods applied to the category fluency test and other tests of semantic memory. Front. Psychol. 14, 1152574. https://doi.org/10.3389/fpsyg.2023.1152574.

De Marco, M., Wright, L.M., Makovac, E., 2025. Item-level analysis of Category Fluency test performance: A systematic review and meta-analysis of studies of normal and neurologically-abnormal ageing. Neuropsychol. Rev. https://doi.org/10.1007/s11065-024-09657-z.

Dubois, B., Slachevsky, A., Litvan, I., Pillon, B., 2000. The FAB: A Frontal Assessment Battery at bedside. Neurology 55, 1621–1626. https://doi.org/10.1212/wnl.55.11.1621.

Elias, M.F., Beiser, A., Wolf, P.A., Au, R., White, R.F., D'Agostino, R.B., 2000. The preclinical phase of Alzheimer disease: A 22-year prospective study of the Framingham Cohort. Arch. Neurol. 57, 808–813. https://doi.org/10.1001/archneur.57.6.808.

Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc. Natl. Acad. Sci. U. S. A. 97, 11050–11055. https://doi.org/10.1073/pnas.200033797.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355. https://doi.org/10.1016/s0896-6273(02)00569-x.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A. M., 2004. Automatically parcellating the human cerebral cortex. Cereb. Cortex 14, 11–22. https://doi.org/10.1093/cercor/bhg087.

Forbes-McKay, K.E., Ellis, A.W., Shanks, M.F., Venneri, A., 2005. The age of acquisition of words produced in a semantic fluency task can reliably differentiate normal from pathological age related cognitive decline. Neuropsychologia 43, 1625–1632. https://doi.org/10.1016/j.neuropsychologia.2005.01.008.

Giovannetti, T., Lamar, M., Cloud, B.S., Swenson, R., Fein, D., Kaplan, E., Libon, D.J., 2001. Different underlying mechanisms for deficits in concept formation in dementia. Arch. Clin. Neuropsychol. 16, 547–560. https://doi.org/10.1016/s0887-6177(00)00066-4.

Gruenewald, P.J., Lockhead, G.R., 1980. The free recall of category examples. J. Exp. Psychol. Hum. Learn. Mem. 6, 225–240. https://doi.org/10.1037/0278-7393.6.3.225.

Herlin, B., Navarro, V., Dupont, S., 2021. The temporal pole: From anatomy to function-A literature appraisal. J. Chem. Neuroanat. 113, 101925. https://doi.org/10.1016/j.jchemneu.2021.101925.

Hétu, S., Grégoire, M., Saimpont, A., Coll, M.P., Eugène, F., Michon, P.E., Jackson, P.L., 2013. The neural network of motor imagery: An ALE meta-analysis. Neurosci. Biobehav. Rev. 37, 930–949. https://doi.org/10.1016/j.neubiorev.2013.03.017.

Hickok, G., 2009. The functional neuroanatomy of language. Phys. Life Rev. 6, 121–143. https://doi.org/10.1016/j.plrev.2009.06.001.

Hirni, D.I., Kivisaari, S.L., Krumm, S., Monsch, A.U., Berres, M., Oeksuez, F., Reinhardt, J., Ulmer, S., Kressig, R.W., Stippich, C., Taylor, K.I., 2016. Neuropsychological markers of medial perirhinal and entorhinal cortex functioning are impaired twelve years preceding diagnosis of Alzheimer's dementia. J. Alzheimers Dis. 52, 573–580. https://doi.org/10.3233/jad-150158.

Hirnstein, M., Westerhausen, R., Hugdahl, K., 2013. The right planum temporale is involved in stimulus-driven, auditory attention – Evidence from transcranial magnetic stimulation. PLoS One 8, e57316. https://doi.org/10.1371/journal.pone.0057316.

Huth, A.G., de Heer, W.A., Griffiths, T.L., Theunissen, F.E., Gallant, J.L., 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532, 453–458. https://doi.org/10.1038/nature17637.

Hyatt, C.S., Owens, M.M., Crowe, M.L., Carter, N.T., Lynam, D.R., Miller, J.D., 2020. The quandary of covarying: A brief review and empirical examination of covariate use in structural neuroimaging studies on psychological variables. Neuroimage 205, 116225. https://doi.org/10.1016/j.neuroimage.2019.116225.

Igarashi, K.M., 2023. Entorhinal cortex dysfunction in Alzheimer's disease. Trends Neurosci. 46, 124–136. https://doi.org/10.1016/j.tins.2022.11.006.

Jackson, R.J., 2021. The neural correlates of semantic control revisited. Neuroimage 224, 117444. https://doi.org/10.1016/j.neuroimage.2020.117444.

Jacobs, D.M., Sano, M., Dooneief, G., Marder, K., Bell, K.L., Stern, Y., 1995. Neuropsychological detection and characterization of preclinical Alzheimer's disease. Neurology 45, 957–962. https://doi.org/10.1212/wnl.45.5.957.

Keren-Happuch, E., Chen, S.H.A., Ho, M.H.R., Desmond, J.E., 2014. A meta-analysis of cerebellar contributions to higher cognition from PET and fMRI studies. Hum. Brain Mapp. 35, 593–615. https://doi.org/10.1002/hbm.22194.

Klein, L.A., Buchanan, J.A., 2009. Psychometric properties of the Pyramids and Palm Trees Test. J. Clin. Exp. Neuropsychol. 31, 803–808. https://doi.org/10.1080/13803390802508926.

Krumm, S., Kivisaari, S.L., Probst, A., Monsch, A.U., Reinhardt, J., Ulmer, S., Stippich, C., Kressig, R.W., Taylor, K.I., 2016. Cortical thinning of parahippocampal subregions in very early Alzheimer's disease. Neurobiol. Aging 38, 188–196. https://doi.org/10.1016/j.neurobiolaging.2015.11.001.

Lambon Ralph, M.A., Jefferies, E., Patterson, K., Rogers, T.T., 2017. The neural and computational bases of semantic cognition. Nat. Rev. Neurosci. 18, 42–55. https://doi.org/10.1038/nrn.2016.150.

Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P.V., Nickerson, D., Mikiten, S.A., Fox, P.T., 2000. Automated Talairach atlas labels for functional brain mapping. Hum. Brain Mapp. 10, 120–131. https://doi.org/10.1002/1097-0193(200007)10:3<120::aid-hbm30>3.0.co;2-8.

Leacock, C., Chorodow, M., 1998. Combining local context with WordNet similarity for word sense identification. In Christiane Fellbaum (ed). WordNet: A lexical reference system and its application. MIT Press, Cambridge, MA, USA.

Liu, L., Liu, D., Guo, T., Schwieter, J.W., Liu, H., 2023. The right superior temporal gyrus plays a role in semantic-rule learning: Evidence supporting a reinforcement learning model. Neuroimage 282, 120393. https://doi.org/10.1016/j.neuroimage.2023.120393.

Lundin, N.B., Brown, J.W., Johns, B.T., Jones, M.N., Purcell, J.R., Hetrick, W.P., O'Donnell, B.F., Todd, P.M., 2023. Neural evidence of switch processes during semantic and phonetic foraging in human memory. Proc. Natl. Acad. Sci. U. S. A. 120, e2312462120. https://doi.org/10.1073/pnas.2312462120.

Markesbery, W.R., 2010. Neuropathologic alterations in mild cognitive impairment: A review. J. Alzheimers Dis. 19, 221–228. https://doi.org/10.3233/jad-2010-1220.

Menenti, L., Petersson, K.M., Scheeringa, R., Hagoort, P., 2009. When elephants fly: differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. J. Cogn. Neurosci. 21, 2358–2368. https://doi.org/10.1162/jocn.2008.21163.

Merz, Z.C., Van Patten, R., Hurless, N., Grant, A., McGrath, A.B., 2021. Furthering the understanding of Wechsler Adult Intelligence Scale-Fourth Edition factor structure in a clinical sample. Appl. Neuropsychol. Adult 28, 12–23. https://doi.org/10.1080/23279095.2019.1585351.

Mishkin, M., Suzuki, W.A., Gadian, D.G., Vargha-Khadem, F., 1997. Hierarchical organization of cognitive memory. Philos. Trans. R. Soc. Lond. B Biol. Sci. 352, 1461–1467. https://doi.org/10.1098/rstb.1997.0132.

Mueller, K.D., Du, L., Bruno, D., Betthauser, T., Christian, B., Johnson, S., Hermann, B., Koscik, R.L., 2023. Item-level story recall predictors of amyloid-beta in late middle-aged adults at increased risk for Alzheimer's disease. Front. Psychol. 13, 908651. https://doi.org/10.3389/fpsyg.2022.908651.

Payton, N.M., Rizzuto, D., Fratiglioni, L., Kivipelto, M., Bäckman, L., Laukka, E.J., 2020. Combining cognitive markers to identify individuals at increased dementia risk: Influence of modifying factors and time to diagnosis. J. Int. Neuropsychol. Soc. 26, 785–797. https://doi.org/10.1017/s1355617720000272.

Pérez-Cordón, A., Monté-Rubio, G., Sanabria, A., Rodriguez-Gomez, O., Valero, S., Abdelnour, C., Marquié, M., Espinosa, A., Ortega, G., Hernandez, I., Rosende-Roca, M., Vargas, L., Mauleón, A., Gil, S., Tartari, J.P., Lomeña, F., Campos, F., Vivas, A., Gomez-Chiari, M., Benaque, A., Ruiz, A., Tárraga, L., Boada, M., Alegret, M., FACEHBI Group, 2020. Subtle executive deficits are associated with higher brain amyloid burden and lower cortical volume in subjective cognitive decline: The FACEHBI cohort. Sci. Rep. 10, 17721. https://doi.org/10.1038/s41598-020-74704-7.

Petersen, R.C., Lopez, O., Armstrong, M.J., Getchius, T.S.D., Ganguli, M., Gloss, D., Gronseth, G.S., Marson, D., Pringsheim, T., Day, G.S., Sager, M., Stevens, J., Rae-Grant, A., 2018. Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. Neurology 90, 126–135. https://doi.org/10.1212/WNL.0000000000004826.

Sachdev, P.S., Blacker, D., Blazer, D.G., Ganguli, M., Jeste, D.V., Paulsen, J.S., Petersen, R.C., 2014. Classifying neurocognitive disorders: The DSM-5 approach. Nat. Rev. Neurol. 10, 634–642. https://doi.org/10.1038/nrneurol.2014.181.

Shapleske, J., Rossell, S.L., Woodruff, P.W.R., David, A.S., 1999. The planum temporale: A systematic, quantitative review of its structural, functional and clinical significance. Brain Res. Rev. 29, 26–49. https://doi.org/10.1016/s0165-0173(98)00047-2.

Snodgrass, J.G., Vanderwart, M., 1980. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. J. Exp. Psychol. Hum. Learn 6, 174–215. https://doi.org/10.1037/0278-7393.6.2.174.

Stokholm, J., Vogel, A., Gade, A., Waldemar, G., 2006. Heterogeneity in executive impairment in patients with very mild Alzheimer's disease. Dement. Geriatr. Cogn. Disord. 22, 54–59. https://doi.org/10.1159/000093262.

Tesink, C.M.J.Y., Buitelaar, J.K., Petersson, K.M., van der Gaag, R.J., Kan, C.C., Tendolkar, I., Hagoort, P., 2009a. Neural correlates of pragmatic language comprehension in autism spectrum disorders. Brain 132, 1941–1952. https://doi.org/10.1093/brain/awp103.

Tesink, C.M., Petersson, K.M., van Berkum, J.J., van den Brink, D., Buitelaar, J.K., Hagoort, P., 2009b. Unification of speaker and meaning in language comprehension: An FMRI study. J. Cogn. Neurosci. 21, 2085–2099. https://doi.org/10.1162/jocn.2008.21161.

Thompson, C.G., Kim, R.S., Aloe, A.M., Becker, B.J., 2017. Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. Basic Appl. Soc. Psychol. 39, 81–90. https://doi.org/10.1080/01973533.2016.1277529.

Thompson-Schill, S.L., Aguirre, G.K., D'Esposito, M., Farah, M.J., 1999. A neural basis for category and modality specificity of semantic knowledge. Neuropsychologia 37, 671–676. https://doi.org/10.1016/S0028-3932(98)00126-2.

Turker, S., Kuhnke, P., Eickhoff, S.B., Caspers, S., Hartwigsen, G., 2023. Cortical, subcortical, and cerebellar contributions to language processing: A meta-analytic review of 403 neuroimaging experiments. Psychol. Bull. 149, 699–723. https://doi.org/10.1037/bul0000403.

Venneri, A., Jahn-Carta, C., De Marco, M., Quaranta, D., Marra, C., 2018. Diagnostic and prognostic role of semantic processing in preclinical Alzheimer's disease. Biomark. Med. 12, 637–651. https://doi.org/10.2217/bmm-2017-0324.

Venneri, A., Mitolo, M., De Marco, M., 2016. Paradigm shift: Semantic memory decline as a biomarker of preclinical Alzheimer's disease. Biomark. Med. 10, 5–8. https://doi.org/10.2217/bmm.15.53.

Vonk, J.M.J., Flores, R.J., Rosado, D., Qian, C., Cabo, R., Habegger, J., Louie, K., Allocco, E., Brickman, A.M., Manly, J.J., 2019. Semantic network function captured by word frequency in nondemented APOE ε4 carriers. Neuropsychology 33, 256–262. https://doi.org/10.1037/neu0000508.

Wechsler, D., 2008. Wechsler Adult Intelligence Scale–Fourth Edition technical and interpretive manual. Pearson, San Antonio, TX.

Whiteside, D.M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M.R., Roper, B., 2016. Verbal fluency: Language or executive function measure? Appl. Neuropsychol. Adult 23, 29–34. https://doi.org/10.1080/23279095.2015.1004574.

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L.O., Nordberg, A., Bäckman, L., Albert, M., Almkvist, O., Arai, H., Basun, H., Blennow, K., de Leon, M., DeCarli, C., Erkinjuntti, T., Giacobini, E., Graff, C., Hardy, J., Jack, C., Jorm, A., Ritchie, K., van Duijn, C., Visser, P., Petersen, R.C., 2004. Mild cognitive impairment–beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. J. Intern. Med. 256, 240–246. https://doi.org/10.1111/j.1365-2796.2004.01380.x.

Xie, L., Das, S.R., Wisse, L.E.M., Ittyerah, R., Yushkevich, P.A., Wolk, D.A., for the Alzheimer's Disease Neuroimaging Initiative, 2018. Early Tau burden correlates with higher rate of atrophy in transentorhinal cortex. J. Alzheimers Dis. 62, 85–92. https://doi.org/10.3233/jad-170945.

Zalonis, I., Christidi, F., Paraskevas, G., Zabelis, T., Evdokimidis, I., Kararizou, E., 2012. Can executive cognitive measures differentiate between patients with spinal- and bulbar-onset amyotrophic lateral sclerosis? Arch. Clin. Neuropsychol. 27, 348–354. https://doi.org/10.1093/arclin/acs031.