

Journal of Cognitive Systems and Continuity

VOLUME 11 | ISSUE 1 | MARCH 2019

Continuity Markers in Late-Stage Cognitive Decline

A Framework for Longitudinal Behavioural Assessment

Prof. Jane Ridley¹ · Dr Eleanor Marsh² · Dr Simon Beck³

¹ Science Directorate, The Society, Shanghai

² Continuity Research, Society London Office

³ Behavioural Systems, Society London Office

Article type	Original Research
Received	4 October 2018
Accepted	19 February 2019
Published	18 March 2019

ISSN 2057-7421 (Print) | ISSN 2057-743X (Online) | DOI: 10.21941/JCSC.2019.0318

Abstract

Background: Conventional dementia assessment is dominated by measures of episodic memory, orientation and executive function. Such measures are clinically useful but do not directly quantify the persistent behavioural organisation through which relatives and clinicians often continue to recognise an individual. **Objective:** To define a set of observable continuity markers, develop a conservative within-person Continuity Index (CI), and test the feasibility and reliability of longitudinal scoring in late-stage cognitive decline. **Methods:** Twenty-two participants with probable Alzheimer's disease or mixed Alzheimer's/vascular dementia were assessed at baseline and at 3-month intervals for 18 months. Fourteen trained observers, blinded to prior ratings, scored seven domains from standardised interviews and naturalistic task recordings. The CI combined domain stability scores with a penalty for abrupt within-person variance. **Results:** Mean MMSE declined from 18.7 to 10.4 (-44%), whereas mean CI declined from 0.84 to 0.66 (-21%). Inter-rater reliability was acceptable to good in six domains (ICC 0.74-0.85) and lower for narrative organisation (ICC 0.68). Moral consistency and preference hierarchy showed the smallest mean decline. **Conclusions:** Behavioural continuity can be operationalised with moderate reliability and appears to decline more slowly than conventional global cognitive performance. The CI is not a measure of consciousness, legal competence or metaphysical identity; it is a descriptive longitudinal instrument requiring replication in larger and more diverse cohorts.

Keywords: dementia; behavioural continuity; longitudinal assessment; identity; autobiographical memory; observational measurement

1. Introduction

Clinical descriptions of neurodegenerative disease necessarily attend to loss: impaired recent memory, disorientation, language disturbance, declining executive control and reduced functional independence. The Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA) remain useful screening instruments, but neither was designed to represent the totality of an individual's recognisable behavioural organisation. In practice, relatives often report that a person remains recognisably themselves through characteristic humour, conversational rhythm, ethical reactions, enduring dislikes, or a familiar way of approaching an everyday problem, even after substantial deterioration in episodic recall.

This distinction is not merely semantic. Autobiographical memory contributes strongly to self-knowledge and narrative identity, yet memory is neither unitary nor uniformly affected across dementia syndromes. Episodic detail may be lost while semantic knowledge, affective associations, procedural habits and interpersonal dispositions persist to different degrees. A framework concerned only with retrieved facts may therefore underdescribe what carers mean when they say that a person is 'still there', while an unstructured appeal to personhood risks importing philosophical conclusions that cannot be supported by observation alone.

We use the term behavioural continuity to denote persistence in the organisation of observable responses across repeated encounters. The construct is deliberately narrower than personal identity. It does not presume that stable behaviour establishes subjective continuity, nor that behavioural discontinuity entails its absence. The present study asks a more limited empirical question: can recognisable patterns be described consistently enough to support longitudinal measurement?

1.1 Aims and scope

The study had three aims: (i) to define a provisional set of continuity-marker domains that can be observed in late-stage cognitive decline; (ii) to evaluate whether trained observers can score those domains with acceptable agreement; and (iii) to examine whether the resulting index changes at a rate distinguishable from conventional cognitive screening. The framework is intended for research use and within-person comparison only.

2. Conceptual background

2.1 Memory, self-representation and recognisability

Philosophical accounts have often treated memory continuity as central to personal identity, but contemporary cognitive models distribute self-representation across autobiographical memory, semantic knowledge, affective

valuation, social cognition and embodied habits. In Alzheimer's disease, autobiographical recall commonly becomes overgeneral and decontextualised. Nevertheless, emotionally salient preferences, familiar interpersonal responses and well-learned behavioural routines may remain available after detailed episodic reconstruction is impaired.

Recognisability is not equivalent to identity. It is an observer-dependent judgement shaped by relationship history, expectation and cultural interpretation. For that reason, the present framework avoids asking observers whether a participant 'seems like the same person'. Instead, observers score narrowly defined features, such as the stability of preferred vocabulary, ordering of narrative material, relative weighting of competing values, or recurrent strategy selection in a standardised task.

2.2 Existing assessment approaches

Global cognitive instruments provide efficient estimates of impairment but necessarily compress heterogeneous functions into a single score. Neuropsychiatric inventories and functional scales capture behavioural symptoms and disability, yet they generally measure frequency or severity rather than preservation of an individual's characteristic pattern. Informant-based measures of self and personality provide useful precedent, but are vulnerable to retrospective reconstruction and may not separate stable traits from relationship-specific expectations.

The proposed CI is therefore complementary. It seeks neither diagnostic sensitivity nor between-person ranking. Its intended use is to describe the persistence and reorganisation of an individual's behavioural pattern across time.

3. Development of continuity-marker domains

Candidate domains were generated from structured interviews with clinicians and family carers, review of recorded dementia assessments, and theoretical analysis of behavioural features expected to show within-person stability. Domains were retained when they met four criteria: observability in a 30-45 minute session; plausible independence from simple episodic recall; relevance across different educational backgrounds; and capacity for repeated scoring without requiring knowledge of the participant's diagnosis or prior score.

Domain	Operational definition	Illustrative observation
Linguistic cadence	Characteristic vocabulary, syntax, pacing and prosody independent of factual accuracy.	Repeated preference for compressed phrasing and delayed response onset.
Narrative organisation	Typical sequencing, framing and closure of spoken accounts.	Begins with contextual qualification before addressing the event.
Humour profile	Stable triggers, form and social function of humour.	Uses self-deprecation to reduce interpersonal tension.
Emotional weighting	Relative salience assigned to emotional cues and consequences.	Responds more strongly to perceived unfairness than inconvenience.
Preference hierarchy	Consistency in ranking familiar options and values.	Chooses privacy over convenience across varied scenarios.
Moral consistency	Stability of ethical priorities and justificatory style.	Prioritises harm avoidance even when rules permit another action.
Problem-solving strategy	Characteristic approach to novel or mildly unfamiliar tasks.	Attempts verbal decomposition before trial-and-error action.

Social recognition was considered as an eighth domain but excluded from the principal index. Recognition performance was too directly dependent on memory, sensory acuity and cueing conditions, and its inclusion risked conflating continuity with retrieval success.

4. Continuity Index

Each domain score, $S_i(t)$, ranges from 0 to 1 and represents similarity between behaviour observed at time t and the participant's baseline profile. Scores are based on rubric anchors rather than raw frequency. A score of 1.0 indicates

strong preservation of the characteristic pattern; 0.5 indicates partial preservation with substantial reorganisation; and 0 indicates absence of the baseline pattern or emergence of a consistently incompatible pattern.

The preliminary Continuity Index is defined as:

$$CI(t) = \sum w_i S_i(t) - \lambda V(t)$$

where w_i is the provisional weight assigned to domain i ($\sum w_i = 1$), $V(t)$ is the mean absolute divergence between adjacent assessments across domains, and λ is an instability coefficient fixed at 0.10 for this pilot study. The variance penalty reduces the index where apparent stability is produced by alternating or internally inconsistent ratings. CI values are bounded to the interval [0,1].

Domain	Weight
Linguistic cadence	0.15
Narrative organisation	0.13
Humour profile	0.12
Emotional weighting	0.14
Preference hierarchy	0.17
Moral consistency	0.17
Problem-solving strategy	0.12

Weights were set by expert consensus before outcome analysis. Higher weights were assigned to preference hierarchy and moral consistency because preliminary work suggested comparatively good discriminability and lower dependence on momentary task performance. These values should not be treated as optimised coefficients.

5. Methods

5.1 Design and participants

This was an 18-month prospective observational feasibility study conducted through two memory services in London and Oxford. Twenty-two participants were recruited between January and March 2017. Inclusion criteria were: age 60-90 years; clinical diagnosis of probable Alzheimer's disease or mixed Alzheimer's/vascular dementia; baseline MMSE 10-24; availability of an informant with at least weekly contact; and sufficient hearing and English comprehension to participate in recorded assessment. Exclusion criteria were major psychiatric disorder predating dementia, acute delirium, unstable neurological illness, or sensory impairment preventing meaningful participation.

The final cohort comprised 14 women and 8 men; mean age was 76.8 years (SD 6.9), mean education 12.1 years (SD 3.4), and mean time since diagnosis 3.6 years (SD 1.8). Seventeen participants had probable Alzheimer's disease and five had mixed Alzheimer's/vascular dementia.

5.2 Procedure

Participants were assessed at baseline and at 3, 6, 9, 12, 15 and 18 months. Sessions included the MMSE, a 12-minute semi-structured conversation, a picture-supported narrative task, two preference-ranking tasks, three brief moral scenarios and a practical problem-solving task. Sessions were video-recorded using a fixed camera and lapel microphone. Assessors used standard prompts and were instructed not to correct factual errors unless required for safety.

At baseline, an informant interview established examples of the participant's characteristic speech, humour, preferences, value priorities and problem-solving habits during the five years preceding diagnosis. These data were used to construct a concise baseline profile. Observers scoring follow-up sessions had access to the baseline profile and baseline recording but were blinded to diagnosis subtype, MMSE score, assessment date and all prior follow-up ratings.

5.3 Observer training and scoring

Fourteen observers (six clinical psychologists, four speech and language therapists, two occupational therapists and two research clinicians) completed a 30-hour training programme. Training included rubric instruction, calibration on eight recordings not included in the study, and discussion of disagreements. Observers were required to reach a two-way random-effects intraclass correlation coefficient (ICC) of at least 0.70 on calibration material before contributing study ratings.

Each recording was scored independently by two observers. A third observer rated a randomly selected 25% sample to estimate reliability under variable rater pairing. Ratings were entered before observers could view any other score.

5.4 Statistical analysis

Analyses were prespecified as exploratory. Inter-rater reliability was estimated using ICC(2,1) with 95% confidence intervals. Longitudinal change was summarised as percentage change from baseline and modelled using linear mixed-effects regression with random intercepts for participant and fixed effects of time. Because the cohort was small and the primary objective was feasibility, emphasis was placed on effect estimates and confidence intervals rather than dichotomous significance testing. Bootstrap confidence intervals (10,000 resamples) were used for mean percentage change. Missing domain observations were handled by renormalising weights when no more than one domain was missing; assessments missing two or more domains were excluded from the principal CI analysis.

5.5 Ethics

Written informed consent was obtained from participants judged to have capacity. Where capacity was absent or subsequently lost, participation proceeded only with consultee agreement and ongoing assent, in accordance with the Mental Capacity Act 2005. The protocol was approved by the Society Research Ethics Committee (Ref: SRE-2016-214). Participation had no effect on clinical care.

6. Results

6.1 Completion and missing data

Nineteen participants completed the 18-month assessment. Two died from causes unrelated to the study and one withdrew after relocation. Of 154 scheduled sessions, 145 (94.2%) were completed. Seven sessions contained one unscorable domain because of fatigue or recording failure; no session had more than two missing domains. Principal CI analyses included 141 sessions.

6.2 Inter-rater reliability

Domain	ICC (2,1)	95% CI
Linguistic cadence	0.78	0.66-0.87
Narrative organisation	0.68	0.54-0.79
Humour profile	0.76	0.63-0.84
Emotional weighting	0.81	0.70-0.89
Preference hierarchy	0.85	0.76-0.92
Moral consistency	0.79	0.68-0.87
Problem-solving strategy	0.73	0.59-0.83
Composite CI	0.86	0.78-0.91

Agreement was highest for preference hierarchy and lowest for narrative organisation. Review of discordant narrative ratings indicated that observers differed in whether factual incompleteness should reduce organisation scores when the participant retained a characteristic framing and sequence. The rubric was amended after data lock

6.3 Longitudinal trajectories

Measure	Baseline	18 months	Relative change	Bootstrap 95% CI
MMSE	18.7 (3.4)	10.4 (4.1)	-44%	-51% to -36%
Continuity Index	0.84 (0.06)	0.66 (0.11)	-21%	-26% to -16%
Narrative organisation	0.81	0.60	-26%	-33% to -19%
Moral consistency	0.88	0.81	-8%	-13% to -3%
Preference hierarchy	0.91	0.84	-8%	-12% to -4%
Problem-solving strategy	0.80	0.66	-18%	-25% to -11%

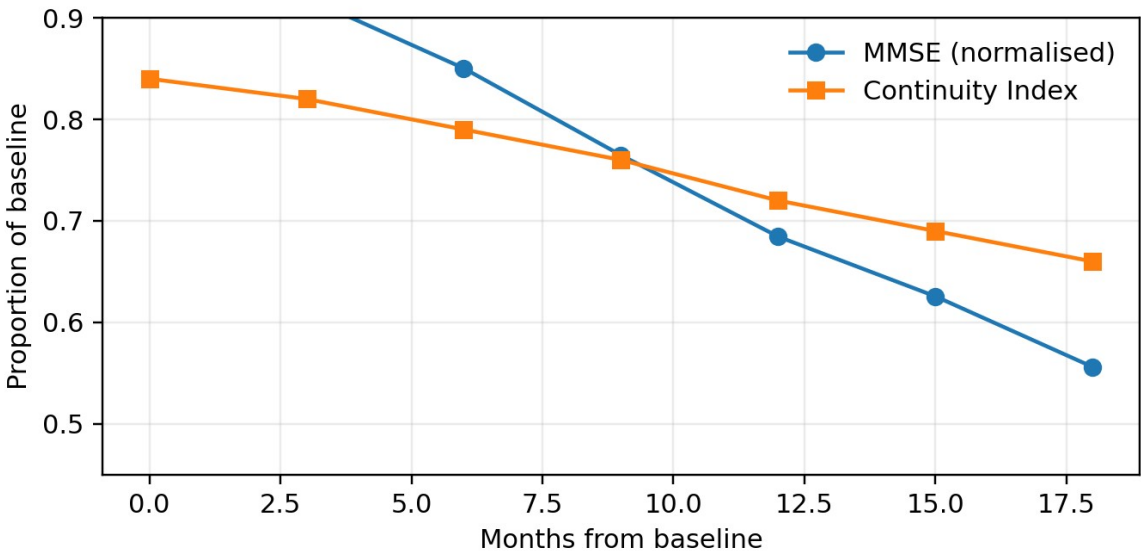
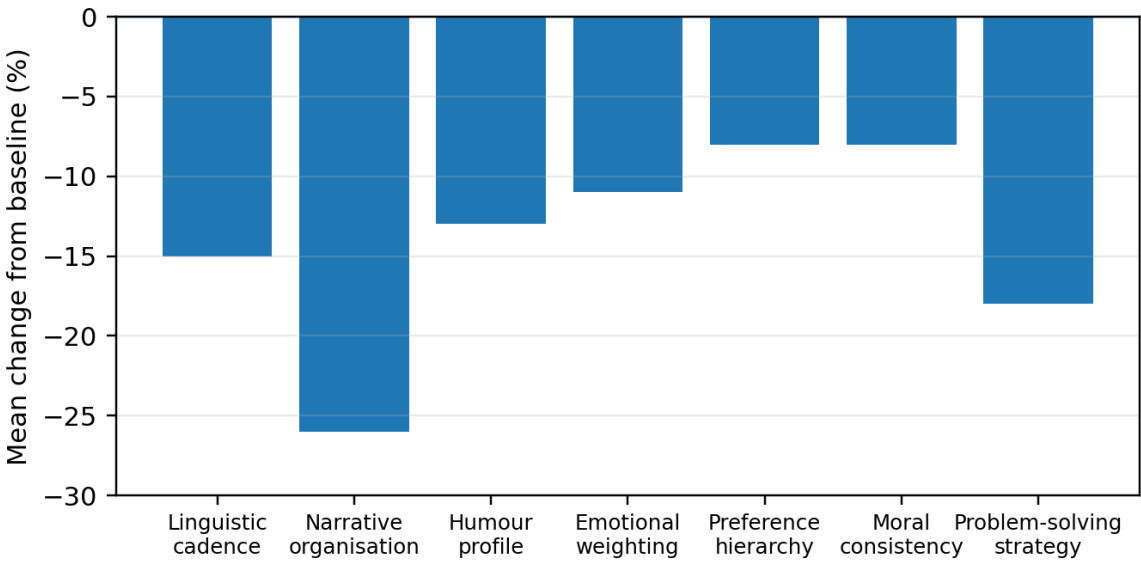


Figure 1. Mean normalised MMSE and Continuity Index trajectories over 18 months. Error bars omitted for visual clarity; analyses used participant-level mixed models.



The estimated monthly decline was -0.47 MMSE points (95% CI -0.56 to -0.39) and -0.010 CI units (95% CI -0.013 to -0.007). When both outcomes were expressed as proportions of baseline, MMSE declined more rapidly than CI (time-by-measure interaction: -0.012 per month, 95% CI -0.017 to -0.008). Individual trajectories varied substantially. Three participants showed abrupt CI decreases associated with intercurrent illness or transition to residential care, followed by partial recovery at the next assessment.

6.4 Domain-level findings

Preference hierarchy and moral consistency were the most stable domains. Participants frequently retained the relative ordering of familiar choices even when unable to explain when or why those preferences developed. Moral scenario responses became shorter and less elaborated over time but often preserved the participant's characteristic priority, such as avoiding interpersonal harm, respecting promises or resisting perceived coercion.

Narrative organisation declined most markedly. Some participants retained characteristic openings or evaluative comments but lost temporal ordering and closure. Humour was absent in many individual sessions, creating floor effects; however, where humour occurred, its form and interpersonal use were often recognisable. Linguistic cadence generally remained stable until later assessments, although reduced output limited scoring in four participants.

6.5 Relationship with conventional cognition

Baseline CI was moderately associated with MMSE ($r = 0.41$), but change scores were weakly correlated ($r = 0.24$). This pattern is consistent with partial overlap rather than equivalence. The CI was more strongly related to informant-rated recognisability at month 18 (Spearman $\rho = 0.62$) than MMSE was ($\rho = 0.29$). Because informant recognisability was exploratory and not independently validated, this comparison should be interpreted cautiously.

7. Discussion

The pilot findings support the feasibility of measuring behavioural continuity in progressive cognitive decline. Trained observers achieved acceptable reliability in six of seven domains, and the composite CI declined substantially more slowly than MMSE performance over 18 months. The result does not imply that identity is preserved in proportion to the CI, nor that a person with a low CI has lost identity. It indicates only that certain organised behavioural relations can remain observable after marked decline in global cognitive performance.

The strongest stability was observed in preference hierarchy and moral consistency. These domains may depend less on explicit autobiographical retrieval than on long-established affective and evaluative structures. Their persistence accords with clinical observations that people may continue to care about familiar relationships, react characteristically to unfairness, or prefer a known mode of living despite inability to narrate the history of those commitments. Conversely, narrative organisation was comparatively vulnerable, likely because coherent narrative requires memory, language, executive sequencing and monitoring simultaneously.

A notable finding was the sensitivity of CI to contextual disruption. Abrupt declines associated with infection, hospital admission or relocation sometimes partially reversed. This suggests that continuity scores may reflect an interaction between underlying capacity and environmental support. A behaviour that appears absent in an unfamiliar institutional setting may re-emerge in a familiar conversational context. Future versions of the framework should therefore distinguish loss of a pattern from failure to elicit it.

7.1 Interpretation and limits of inference

Behavioural similarity can be produced by several mechanisms: preservation of underlying dispositions, repetition of overlearned routines, environmental cueing, or observer projection. The CI cannot discriminate among these explanations. Nor can it establish consciousness, subjective continuity, legal capacity or moral status. These are separate questions requiring conceptual, legal and clinical analysis beyond the scope of an observational index.

The framework is also intentionally within-person. Comparing a score of 0.70 in one participant with 0.70 in another is not meaningful because baseline profiles differ in richness, observability and contextual dependence. The index is best understood as a structured summary of change relative to a documented baseline.

7.2 Clinical and research relevance

Used cautiously, continuity profiles may improve communication between families and care teams by identifying behaviours and values that remain accessible. They may also help evaluate whether an intervention preserves characteristic engagement rather than merely improving test performance. However, premature clinical adoption would be inappropriate. The current rubric requires substantial training, baseline information may be unavailable, and cultural variation in humour, emotional expression and moral justification has not been tested.

8. Limitations

The study has several important limitations. First, the cohort was small, diagnostically narrow and drawn from two UK services. Second, the baseline profile relied partly on informant report and may reflect idealisation or selective memory. Third, observers had access to baseline recordings, which is necessary for continuity scoring but may inflate apparent stability through expectancy effects. Fourth, domain weights were assigned by consensus rather than estimated independently. Fifth, repeated exposure to tasks may produce familiarity effects. Finally, the study did not compare the CI with detailed personality, neuropsychiatric or functional measures, and it cannot establish incremental predictive validity.

9. Future work

Replication should use larger, multi-centre cohorts spanning different dementia syndromes, languages and cultural settings. The marker set should be subjected to confirmatory factor analysis and measurement-invariance testing. Automated extraction of prosody, lexical choice, gaze, gesture and task strategy may improve reproducibility, although machine-derived features should remain interpretable and should not displace direct behavioural observation.

A second priority is prospective modelling. The present CI describes continuity after comparison with a baseline. Future work should test whether a sufficiently rich baseline profile can predict the form of later behaviour and distinguish gradual adaptation from discontinuous change. Such work may have relevance to assistive systems and neuroprosthetic technologies, but any extension from behavioural prediction to claims about preserved personhood would be conceptually unwarranted.

10. Conclusion

A structured set of continuity markers can describe persistent behavioural organisation during substantial cognitive decline. In this pilot cohort, the Continuity Index showed acceptable inter-rater reliability and declined more slowly than MMSE performance. The framework offers a language for aspects of recognisability that conventional cognitive measures do not capture, while deliberately refusing to equate observable continuity with subjective identity. Larger studies are required before clinical use, and the unresolved distinction between preservation of a person and preservation of the person's observable structure should remain explicit.

Acknowledgements

The authors thank the participants, families, memory-service staff and observer panel. The authors also acknowledge the Society Information Architecture Programme for assistance with data capture and secure storage. No programme personnel contributed to scoring or outcome analysis.

Funding

This work was supported by the Society Research Directorate under the Continuity Methods Initiative (CMI-16-04). The funder had no role in participant recruitment, scoring, statistical analysis or the decision to publish.

Conflicts of interest

The authors declare no financial conflicts of interest. J.R. serves as Chief Scientific Officer of the Society. E.M. directs the Society Continuity Research division and S.B. heads Behavioural Analysis. These institutional roles are disclosed because the proposed framework may inform subsequent Society research programmes.

Data availability

De-identified summary data and the scoring manual are available to qualified researchers on application to the Society Research Ethics Committee. Video data cannot be released because consent did not extend to public distribution and behavioural recordings remain potentially identifiable.

References

1. Addis DR, Tippet LJ. Memory of myself: autobiographical memory and identity in Alzheimer's disease. *Memory*. 2004;12(1):56-74.
2. Baddeley A. The episodic buffer: a new component of working memory? *Trends Cogn Sci*. 2000;4(11):417-423.
3. Caddell LS, Clare L. The impact of dementia on self and identity: a systematic review. *Clin Psychol Rev*. 2010;30(1):113-126.
4. Damasio AR. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace; 1999.
5. El Haj M, Antoine P, Nandrino JL, Kapogiannis D. Autobiographical memory decline in Alzheimer's disease, a theoretical and clinical overview. *Ageing Res Rev*. 2015;23(Pt B):183-192.
6. Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state': a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189-198.
7. Friston K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci*. 2010;11(2):127-138.
8. Hou CE, Miller BL, Kramer JH. Patterns of autobiographical memory loss in dementia. *Int J Geriatr Psychiatry*. 2005;20(9):809-815.
9. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the NIA-AA workgroups. *Alzheimers Dement*. 2011;7(3):263-269.
10. Molloy DW, Alemayehu E, Roberts R. Reliability of a Standardized Mini-Mental State Examination compared with the traditional Mini-Mental State Examination. *Am J Psychiatry*. 1991;148(1):102-105.
11. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695-699.
12. Squire LR. Memory systems of the brain: a brief history and current perspective. *Neurobiol Learn Mem*. 2004;82(3):171-177.
13. Tulving E. Memory and consciousness. *Can Psychol*. 1985;26(1):1-12.
14. Woodberry E, Browne G, Hodges S, Watson P, Kapur N, Woodberry K. The use of a wearable camera improves autobiographical memory in patients with Alzheimer's disease. *Memory*. 2015;23(3):340-349.
15. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191-2194.

Appendix A. Continuity Marker Scoring Anchors

Score	General anchor
1.00	Characteristic pattern clearly present and closely matches baseline across form and function.
0.75	Pattern remains recognisable but is reduced, simplified or intermittently expressed.
0.50	Partial preservation; some baseline features remain but substantial reorganisation is present.
0.25	Only isolated or ambiguous baseline features are observable.
0.00	Baseline pattern not observed, or a stable incompatible pattern predominates.

	is insufficient.
--	------------------

Observers are instructed to score preservation of organisation rather than factual correctness, fluency or social desirability. A domain should be marked not scorable where the task does not provide a genuine opportunity for expression. Absence of humour, for example, is not scored as loss unless the session contains conditions that would ordinarily elicit the participant's established humour profile.

Appendix B. Worked Scoring Examples

B.1 Preserved moral weighting with reduced verbal output

At baseline, Participant 07 consistently rejected actions that exposed another person to avoidable embarrassment, explaining this in extended language. At month 15, the participant could not reconstruct the scenario details but repeatedly selected the lower-harm option and objected to 'making a show of someone'. Moral consistency was scored 0.75: the priority and justificatory direction remained recognisable, although elaboration and contextual control were reduced. Narrative organisation was scored separately and was not protected from decline by the preserved moral response.

B.2 Apparent preference reversal under contextual pressure

Participant 12 had a strong baseline preference for privacy over convenience. During an unfamiliar hospital assessment the participant accepted assistance without objection, producing an initial preference score of 0.25. Review showed that the choice was made after repeated prompting by staff and in the absence of the familiar alternative. At the next home assessment, the established ranking reappeared. The hospital score was retained, but the episode was flagged as context-sensitive and contributed to the variance penalty. This example illustrates why a single discontinuous observation should not be interpreted as permanent change.

B.3 Non-occurrence of humour

Participant 18 used dry corrective humour at baseline, principally when a conversational partner made an overconfident claim. A month-12 recording contained no such opportunity and no humorous response. The domain was marked not scorable rather than zero. At month 15 the assessor deliberately introduced an implausibly certain statement within the standard prompt; the participant responded with a characteristic understated correction, scored 0.75. Elicitation must create an opportunity without scripting the target behaviour.

B.4 Distinguishing factual error from narrative organisation

Participant 03 repeatedly misidentified the location and year of a remembered event but continued to organise accounts by first stating a general rule, then presenting an exception, and finally returning to the rule. Narrative organisation was scored 0.75 despite poor episodic accuracy. The CI records the persistence of the organisational form; factual recall remains represented by conventional cognitive and autobiographical-memory measures.