Precision in neuropsychology

Four challenges when using simplified assumptions

Doctoral thesis

Jacob Stålhammar

Department of Psychiatry and Neurochemistry
Institute of Neuroscience and Physiology
Sahlgrenska Academy at University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2019

Cover illustration by Jacob Stålhammar

Precision in neuropsychology
Four challenges when using
simplified assumptions

jacob.stalhammar@gu.se
jacob@jacobstalhammar.com

ISBN 978-91-7833-420-9 (PRINT)
ISBN 978-91-7833-421-6 (PDF)
http://hdl.handle.net/2077/59540

Typeset by Henrik Robertsson, BrandFactory
Printed in Gothenburg, Sweden 2019
Printed by BrandFactory AB

"It doesn't make any difference how beautiful your guess is.
It doesn't make any difference how smart you are,
who made the guess, or what his name is.
If it disagrees with experiment, it's wrong.
That's all there is to it."
Richard Phillips Feynman (May 11, 1918 – February 15, 1988)

To Mall and Daniel, my dear and curious parents

Precision in neuropsychology

Four challenges when using simplified assumptions

Jacob Stålhammar

Department of Psychiatry and Neurochemistry,
Institute of Neuroscience and Physiology
Sahlgrenska Academy at University of Gothenburg
Gothenburg, Sweden

ABSTRACT

Cognition comprises all thought processes, from perception to memory. Neuropsychological tests are the gold standard (= best way) to measure cognition. However, clinical assessment may at times have to rely on simplified assumptions. This work addresses potential risks of four such assumptions through neuropsychological tests and statistical analysis from: a case report (Paper I); participant data from the Gothenburg Mild Cognitive Impairment study (Papers II, III); and the Swedish Cardio Pulmonary bioImage Study (SCAPIS Pilot, Paper IV). Paper I showed transfer effects from memory training may affect memory tests. Paper II showed that giving free credits for items not administered inflated the scores of those most impaired in the Boston Naming Test (BNT). Paper III showed practice effects could not be ruled out in mild cognitive impairment, and that mean neuropsychological change scores (Δ-scores) described change better than isolated Δ-scores. Paper IV showed that administering neuropsychological tests in Swedish to non-native speakers gave lower results in tests tapping speed and attention, and that vocabulary testing may enhance precision. Conclusion: the four assumptions save time at the cost of precision. In the greatest need for precision, (e.g. for detection of gradual change before manifest loss), considering the above findings will improve assessments.

Keywords: Neuropsychology, practice effects, change scores, mild cognitive impairment, dementia, second language effects, bilingualism

ISBN 978-91-7833-420-9 (PRINT)
ISBN 978-91-7833-421-6 (PDF)

SAMMANFATTNING PÅ SVENSKA

Kognition omfattar allt som har med tanke, minne, språk etc. att göra. Neuropsykologiska test är kognitionsmätningar som bjuds av psykologer. Neuropsykologiska test anses vara “the gold standard” (det bästa sättet) att mäta kognitiv kapacitet, men kraven är olika i forskning och klinik. I en klinisk neuropsykologisk bedömning ingår mer än bara mätning (den kan t o m vara en liten del). Detta arbete fokuserar på mättekniska aspekter med fyra exempel. Exemplen kommer från fyra vetenskapliga arbeten som visar på risker med att på ett förenklat vis använda tidsbesparande antaganden:

Artikel ett, visade hur intensiv träning i minnesteknik gav höga resultat i minnestest. Den första artikeln gav exempel på extrema övningseffekter, utan att personen sett just de testen innan.

Artikel två, visade att tidsbesparing genom att bara ge de svåraste uppgifterna på ett benämningstest (Boston Naming Test, BNT) – men samtidigt ge gratispoäng för ej testade uppgifter – tydligt höjde milt dementa patienters resultat. Gratispoäng gav sämre precision.

Artikel tre, visade fler och större förändringspoäng hos de som led av svårare sjukdom, men också att övningseffekter i enstaka test inte kunde uteslutas. Den tredje artikeln säger att genomsnittet av flera förändringspoäng är säkrare att bedöma än enstaka.

Artikel fyra, visade att test som för en svensk modersmålstalare anses testa “bara” snabbhet, för en person som inte har svenska som modersmål också verkar testa förmågan att benämna något. Språkeffekter påverkade användbarheten hos vanliga snabbhetstest.

Detta arbete visar riskerna med fyra förenklade antaganden: "testsekretess räcker", "testförkortning är riskfritt", "förändringspoäng ger alltid samma slags information", "modersmålseffekter syns bara i verbala deltest". Ingen neuropsykolog är okunnig om dessa risker, tvärtom. Men när behovet av mätnoggrannhet är stort - som vid gradvis förändring, flera år innan manifest sjukdom - då kan precisionen förbättras om man beaktar ovan nämnda fynd.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. Stålhammar, J., Nordlund, A., Wallin, A.
An example of exceptional practice effects in the verbal domain
Neurocase 2015; 21(2):162-8

II. Stålhammar, J., Rydén, I., Nordlund, A., Wallin, A.
Boston Naming Test automatic credits inflate scores of nonaphasic mild dementia patients
J Clin Exp Neuropsychol. 2016; 38(4):381-92

III. Stålhammar, J., Hellström, P., Joas, E., Göthlin, M., Rolstad, S., Eckerström, C., Eckerström, M., Wallin, A.
From slow and stable, to abrupt and variable; the range of mild cognitive impairment-to-dementia neuropsychology change scores
Submitted online 2019-03-03 18:37, ID HAPN-2019-0033.
Applied Neuropsychology: Adult.

IV. Stålhammar, J., Hellström, P., Eckerström, C., Wallin, A.
Neuropsychological test performance of middle aged native and non-native Swedish speakers: No executive advantage
Manuscript

CONTENT

ABBREVIATIONS

Δ	Delta, change
Δ-scores	Change scores
AD	Alzheimer's disease
ADL	Activities of Daily Living
BL	Baseline (first examination or reference point)
BNT	Boston Naming Test (NP test)
CAB	Cognitive Assessment Battery (NP test)
CDR	Clinical Dementia Rating (staging scale)
CDT	Clock Drawing Test (screening test)
CE	Central Executive (theoretical concept)
CNS	Central Nervous System
COWAT	Controlled Oral Word Association Test (NP test)
CSF	CerebroSpinal Fluid (fluid around CNS)
CT	(a.k.a. CAT) Computer Aided Tomography
DSM	Diagnostic and Statistical Manual of Mental Disorders
EXIT	Executive Interview (screening test)
fMRI	Functional MRI (imaging technique)
FU	Follow Up
G-MCI	Gothenburg Mild Cognitive Impairment study
GDS	Global Deterioration Scale (dementia scale)

I-Flex	Investigation of flexibility (screening test)
IQ	Intelligence Quotient
L1	First language, native language
L2	Second language
MCI	Mild Cognitive Impairment
MMSE/MMT	Mini-Mental State Examination/Mini-Mental Test (screening test)
MRI	Magnetic Resonance Imaging (imaging technique)
NP	NeuroPsychological
PASMO	PArallel Serial Mental Operations (NP test)
PET	Positron Emission Tomography (imaging technique)
RAVLT	Rey Auditory Verbal Learning Test (NP test)
RCF	Rey Complex Figure test (NP test)
SCAPIS	Swedish CardioPulmonary Imaging Study
SCI	Subjective Cognitive Impairment (stage level)
SD	Standard Deviation
STEP	Stepwise Comparative Status Examination (screening)
TMT	Trail Making Test (NP test)
VOSP	Visual Object and Space Perception Battery (NP test)
WAIS	Wechsler Adult Intelligence Scale (NP test)
WAIS-III/R	Wechsler Adult Intelligence Scale (versions, NP test)
WLM	Wechsler Logical Memory (NP test)

DEFINITIONS IN SHORT

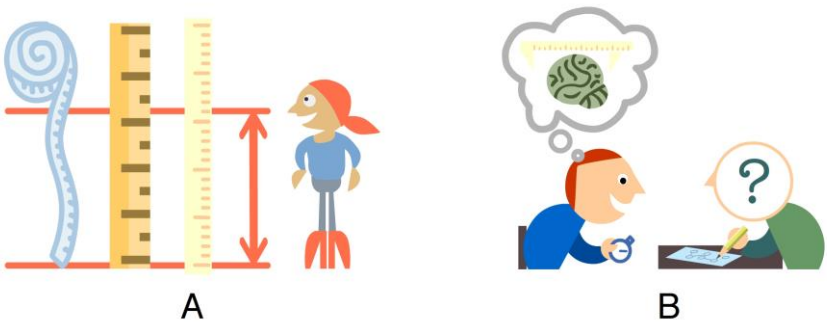
Assumption	Guess or belief held to be true.
Simplified assumption	Simpler guess. May be true in a limited context. Often used to save time. May be good to "get going", yet often performs worse in explanatory contexts.
Cognition	"general term for the processes of thinking" [1]
Neuropsychological test	"sample of behavior obtained under controlled conditions" [2]
Cognitive test	Sampling of cognitive performance, by standardized tests.
Domain	In neuropsychology: a grouping of results from cognitive tests assumed to address similar capacities (e.g. speed/attention, executive attention, learning/memory, verbal, visuo-constructive/spatial).
Activities of Daily Living	In health care: (ADL) a term denoting daily self-care activities (e.g. feeding, grooming, cleaning etc.).
Mild Cognitive Impairment	In health care: (MCI) a stage of objective cognitive impairment – but not at the level of dementia. Neuropsychological test results indicate lower scores compared to age peers, but capacities for ADL are largely intact.
Dementia	"organic loss of intellectual function" [1]

1 INTRODUCTION

1.1 TO MEASURE COGNITION

Perhaps as memory clinic patients suffer acquired impairment, the face of cognitive loss may appear more intimidating. For example, a patient about to lose a driver's license after a failed test may voice a protest, as if saying: "Yes, I saw you measure me, but deep down I know this to be impossible". In some ways this patient would be right, and in some ways the measurement would. The very general aim of writing a thesis on the precision of neuropsychological measurement is to investigate both why a direct measure of cognition is not possible, but also to what degree such an attempt could be informative. The foci are the nature of the source of cognition (the brain), and the methods of measuring cognition.

But first, a primer; measurement is structured observation, expressed in numbers and units. Second, observation can be made in one of two ways: directly, or indirectly (Figure 1). These conditions shaped the history of science and thus the history of measurement.



1 Direct observation (A) allows direct observation of the object of measurement, and direct comparison between measuring instruments. Indirect observation (B) relies on consequences of what is to be measured, and (in the case of neuropsychology) also on effort, motivation, etc.

1.1.1 DIRECT OBSERVATION

Likely since direct observation allows direct comparison, and consequently is easiest to agree upon, our oldest examples of measurement hail from several thousand years BC, with physical examples of units and unit divisions. The first measurement standards likely extended to smaller geographical regions; units were based on common objects (e.g. stones, grains, body parts), and divisions could be thirds, tenths, etc. (e.g. Egyptian cubit [3]). With continuous trade and migration, methods of measurement spread, and the need for wider standardization increased. While several local systems survived in long use, in 1795 a system proposing a natural source for length, with decimal subdivision of units was proposed. The metric system used the 1/10 000 000 distance from the equator to the north pole (through Paris) as its base unit of length, the meter [4]. One tenth (1/10) of a meter cubed became a liter, and a liter of water became a kilogram. Measurement precision was ensured through manufacturing and distribution of physical reference units (metal meters and kilograms). However, for indirect observation things are more complicated.

1.1.2 INDIRECT OBSERVATION

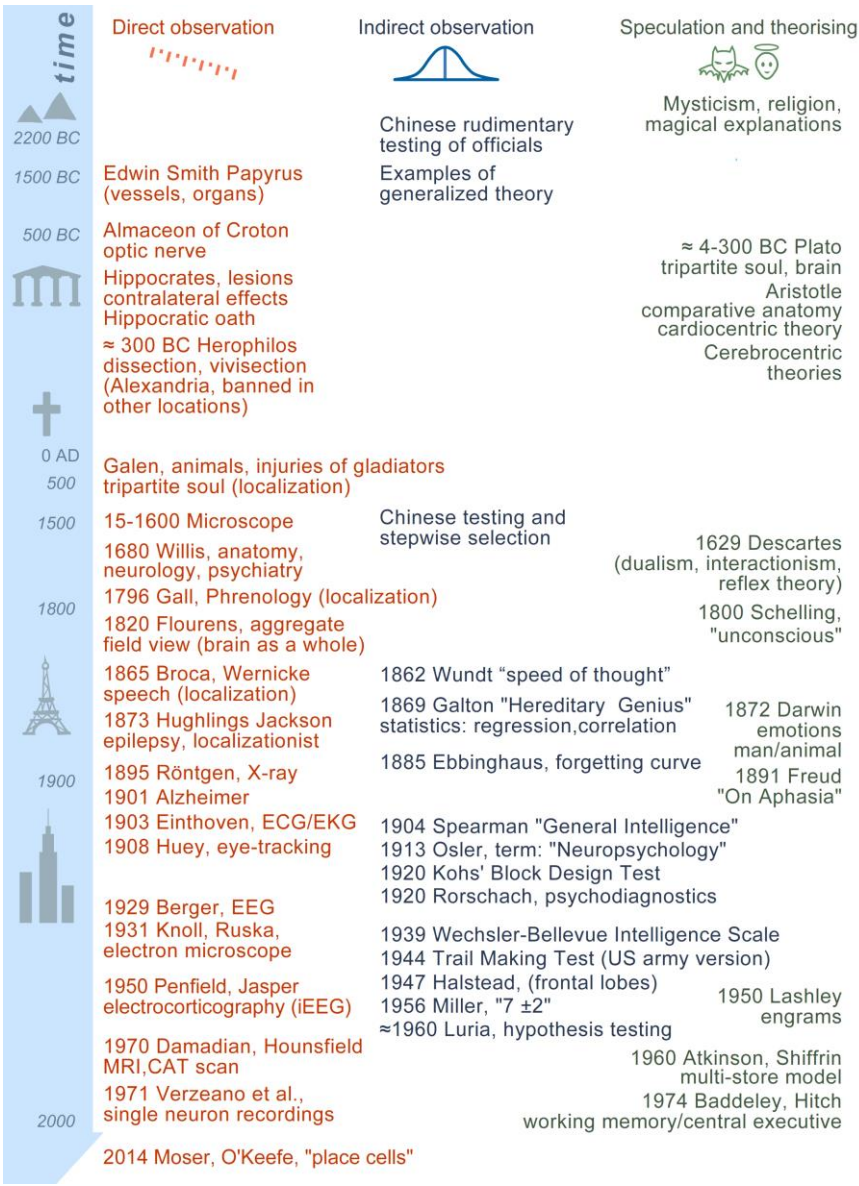
Indirect observation depends on the consequences of something that only might be there. And while religion predated science, and humans have speculated at length on the reasons for their behavior, indirect observation mainly allows comparison of ideas. Thus, to little surprise, at the dawn of science even the very seat of the mind was in debate.

1.1.3 THE SEAT OF THE MIND, EVOLUTION OF SCIENCE

The history of cognition-measurement begins in many ways, but perhaps mostly with speculation on the physical location of cognition. Did we think with our heart? Introspection, e.g. feeling your heart beat faster at the sight of bear, surely indicated the heart as the seat for reaction, and further, without our hearts we died [5]. The brain, on the other hand, was cold to the touch, and experiments of poking it did not evoke pain (e.g. Empedocles, Aristotle [5]). Still, damage to the head resulted in altered behavior, for example in gladiators (e.g. Galen 130 AD, and both brain and spinal injury are mentioned in the Edwin Smith Surgical papyrus (16-1700 BC) [6, 7]).

It's interesting to note that cultures that allowed dissection (e.g. Egypt), even for only religious reasons, still made potentially useful observations. Egyptian mummification, while not a science, practiced that the brain was extracted while the face was preserved, so the soul would find the correct body in the afterlife [8]. Tools and techniques developed to master these delicate operations likely served later Egyptian physicians, who became among the foremost in the ancient world [9]. Around 330 BC the Greek leader Alexander the Great had conquered Egypt, giving his name to the library of Alexandria, where the Corpus Hippocraticum presented brain anatomy [8]. Later, many physicians, e.g. Herophilus (335-280 BC), and Galen of Pergamon (129-199 AD), all studied in Alexandria [9], and theories of the seat of the mind started to point to the brain (even if several operational mechanisms, e.g. pneumatics were proposed). Figure 3 outlines a few milestones, e.g. the microscope in the 15-1600's, Röntgen's X-rays of 1869, Galton's "Hereditary Genius" of 1869, Broca and Wernicke's findings of ca. 1870, Ebbinghaus' forgetting curve of 1885 [1, 10-12]. However, concepts such as "the immortal soul" circulated long, as did speculations on many aspects of human cognition as fundamentally different and superior to that of animals [13].

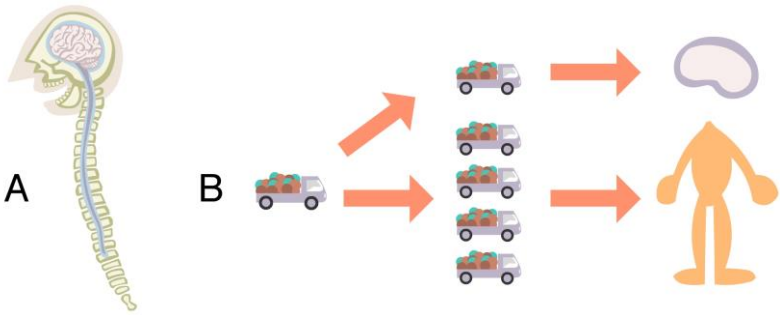
2 Egyptian medical texts are among the oldest [9]. *Accumulation of observation and skills is key, and even methods from religious ceremonies contributed to medical science.*



3 From religious secrets to published science. Brief timeline of types of observation, vs. evolution of neuroscience. The history of neuropsychology, while a mere fraction of that of medicine, begins in structured observation (e.g. testing of Chinese officials 2200 BC [11]*).*

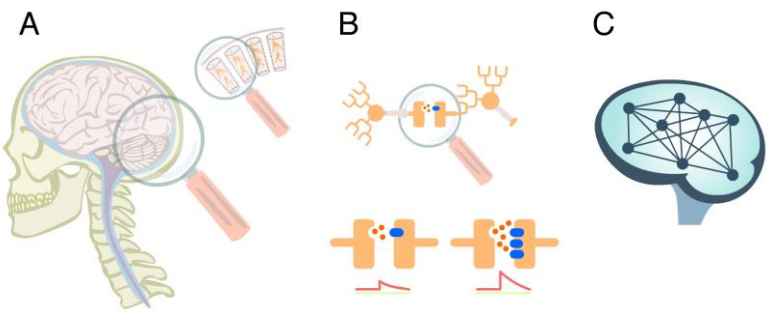
1.2 THE BRAIN

The source of cognition, "the seat of the mind", is the brain. The brain consists of billions of neurons [14]. Yet, more importantly, cognition emerges as a result of the complex systems of networks within networks which neurons help create.



4 A) The brain and the spinal cord make up the central nervous system (CNS). B) The human brain represents ≈ 2% of body weight, yet consumes 20% of the oxygen, and thus calories [15] .

Neurons cluster in several types of functional network units, e.g. ganglia in the autonomic nervous system, cortical columns in the neocortex. Network connections between neurons are not static, but depend on use, synapses are e.g. strengthened from activity patterns by long-term potentiation [1].



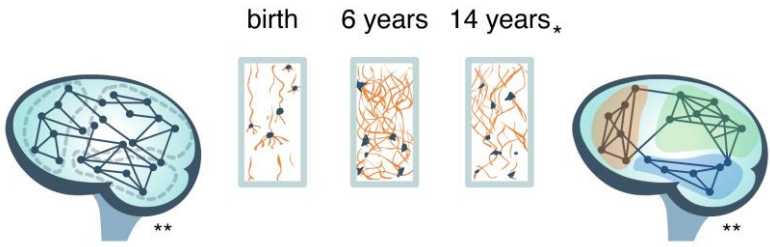
5 A) Brain, cortical columns, B) neurons, synapse connections are strengthened via long-term potentiation, C) the brain is a network of networks.

The neocortex ("gray matter", concentration of neurons) is located on the surface (2-5 mm) of the brain; cerebral white matter, myelinated connec-

tive axons, lies deeper. Due to the innate modularity of the nervous system, many basic functions (e.g. reflexes: sucking, startle, grasp) may be tested at birth [1, 16], yet many neurons are not yet as widely connected, as they eventually will be [1]. In a newborn's first years the number of synapses first grows dramatically (the rate in the macaque peaks at 40.000/s [1] and then declines as fast as 100.000/s [1]). The brain matures via strengthening some connections and pruning others. And as humans age, cortical areas (e.g. frontal cortex, thalamus) have been shown to decrease in size, while areas of cerebral white matter increase, to reach a peak around 50 years, and then decline [14]. The source of cognition is a changing hierarchical structure.

1.2.1 ENERGY CONSERVATION, EVOLUTION, COGNITION

Hierarchical structures are common in nature, and even simulations show that when there is a connection cost, networks evolve to be *both* hierarchical *and* modular [17]. Figure 6 shows brain connection density first increasing, then decreasing. Brain white matter networks grow increasingly modular in adolescent development, affecting frontoparietal areas most and limbic least [18].



6 **Drawing of nerve cells adapted from "Rethinking the Brain"[19]. **Drawing of modular brain adapted from Baum et al. (2017)[18].*

While we cannot directly observe how the brain "makes" cognition, we know sensory information is processed from lower sensory input to higher-order areas in distributed hierarchical systems [1]. And, comparative studies suggest that many neuronal building blocks date back to simpler organisms, or even earlier, e.g. synapse proteins may date back to prokaryotes [20] later passed down to us [16, 21]. Simplified, the brain is a multitude of decision trees built with "use it or lose it" building blocks,

honed by competitive evolution. And, as energy saving gives a competitive edge, reductional processes are central to cognition.

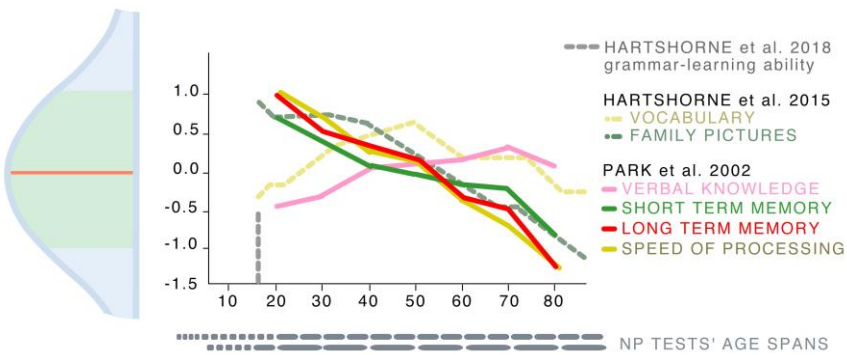
7 Schematic illustration of how perception data (an eye registering a tree) is interpreted/simplified by a network, reducing the amount of data. Much of such a process is automatic and mainly unconscious.

Brain imaging studies show both shrinkage and change in activation pattern e.g. more activation of frontal areas in older individuals [22], unilateral activation in younger, bilateral activation in older [23], and different hippocampal activation [24]) over the normal life span. This suggests that the brain gradually reorganizes e.g. in compensatory scaffolding [25, 26]. Expectedly, cognitive capacities are not stable over time.

8 Schematic illustration of iterative effects of feedback/pruning.

1.2.2 COGNITION OVER THE LIFE SPAN

As the brain undergoes changes, cognitive tests must compare to adequate references. Cognitive measurements in the earlier years of brain development requires both age adequate tasks and reference material from persons within a year or two of the test subject's age, and after the age of 15-25 within 5-10 years [27, 28]. Cognitive capacities continue to change over the lifespan with specific growth curves [29], and different peak performance ages for different domains. Cognitive decline in healthy adults begins in the 20's to 30's [30-32] and cognitive capacities become more variable. In higher age [33, 34] cross sectional visualization of WAIS IV-norms suggested different decline trajectories depending on initial full scale IQ [32] and cognitive capacities show different activation of networks for the same task [24]. Figure 9 shows life span comparison of group-level decline of several domains. Processing speed and capacity for learning decline fastest, while verbal knowledge peaks at around 50 years and then declines [31, 35, 36].



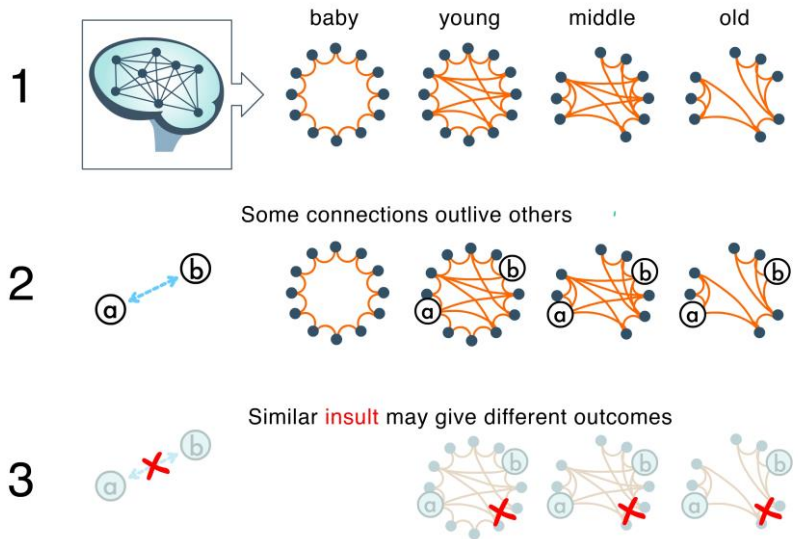
9 Schematic plot of normal cognitive aging (Z-scores, mean 0, SD 1) adapted from [31, 35]. *Dotted divider indicates suggested best-before age of 17.4 years for grammar-learning ability* [36]. *NP-tests age spans approximated from e.g.* [27, 37].

A compounded effect of progressively slower speed but more gradually increasing vocabulary (and likely also heuristic repertoire) may give the impression of a constant-like cognition over the life span. Compensatory recruitment of more frontal areas, and over-activation in older age have been reported [38]. The different relation between patients' observed brain injury and cognitive impairment is sometimes theorized as indicating a brain-reserve (anatomical differences, e.g. hypothetical brain size or synapse count) or cognitive reserve (hypothetically different, e.g. more

efficient "use of the brain")[39]. Previous work has indicated more amyloid pathology in higher educated patients converting to dementia [40]. There is discussion of how to properly define what could constitute compensatory, maintenance and/or reserve principles [26]. However, while compensatory effects may produce the impression of something constant, or even increasing, the brain's resilience to injury changes with age.

1.2.3 BRAIN RESILIENCE

As the brain grows, network topology changes, along with plasticity [1]. Figure 10 shows, simplified, how a similar insult on an un-pruned and a pruned network will produce different results.



10 Principal drawing of why a hypothetical injury may present different results in younger compared to older patients. The insult (red "x" in 3) may not affect the younger network, but will effectively disable linking between "a" and "b" in the older.

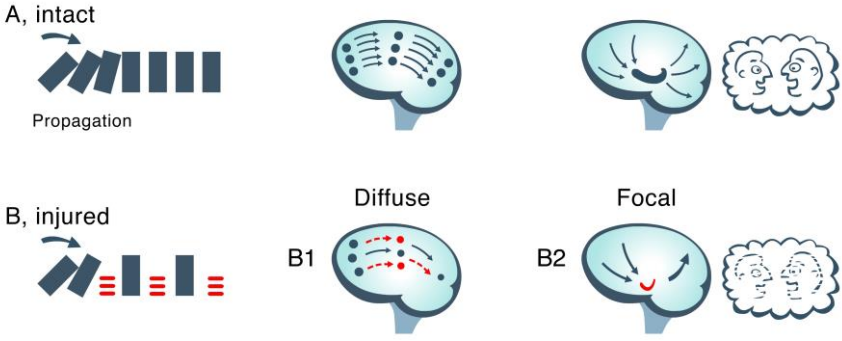
1.2.4 TYPES OF BRAIN DAMAGE, FOCAL AND DIFFUSE

Early knowledge of brain anatomy was gained through case histories, and focal injuries were informative. Well-known examples are e.g. those of "H.M." (Henry Gustav Molaison), an epilepsy patient who lost the capac-

ity to form new episodic memories after surgical removal of two thirds of the hippocampi, or "Tan" (Louis Victor Leborgne), the patient of Broca who lost the capacity for articulated speech after a lesion in what is now known as Broca's area [1].

We now know that vascular injuries may not only produce focal (e.g. Tan) but also more diffuse damage, e.g. hypoperfusion (reduced but not completely blocked blood supply)[41]. As figure 11 illustrates, diffuse injuries may result in general slowness and early/preclinical vascular dementias often present slowness and executive symptoms [42, 43]. Yet, as slowness may affect many neuropsychological examinations, and vascular diseases may also produce focal injury, vascular dementias also often appear heterogeneous [44, 45], although much work has also been done to improve categorization of the many varieties [41].

Figure 11 shows how a diffuse injury may partly and gradually impair signal propagation and thus how a diffuse injury may cause slowness. In contrast, a focal injury may affect "one" function (as in HM, episodic memory). However, studies have also shown how focal damages may affect large networks, depending on location [46].



11 Information propagation depends on network integrity. The signal reaches the next step more efficiently in more intact networks (row A). However, with disruptions in propagation (B), signal may be lost, E.g. small vessel disease may present diffuse injury, and the impression of slowing, (B1). Focal damage (e.g. "HM") may present more distinct symptoms (B2).

1.3 THE HOSPITAL

1.3.1 BRAIN DAMAGE, HOSPITAL ORGANIZATION

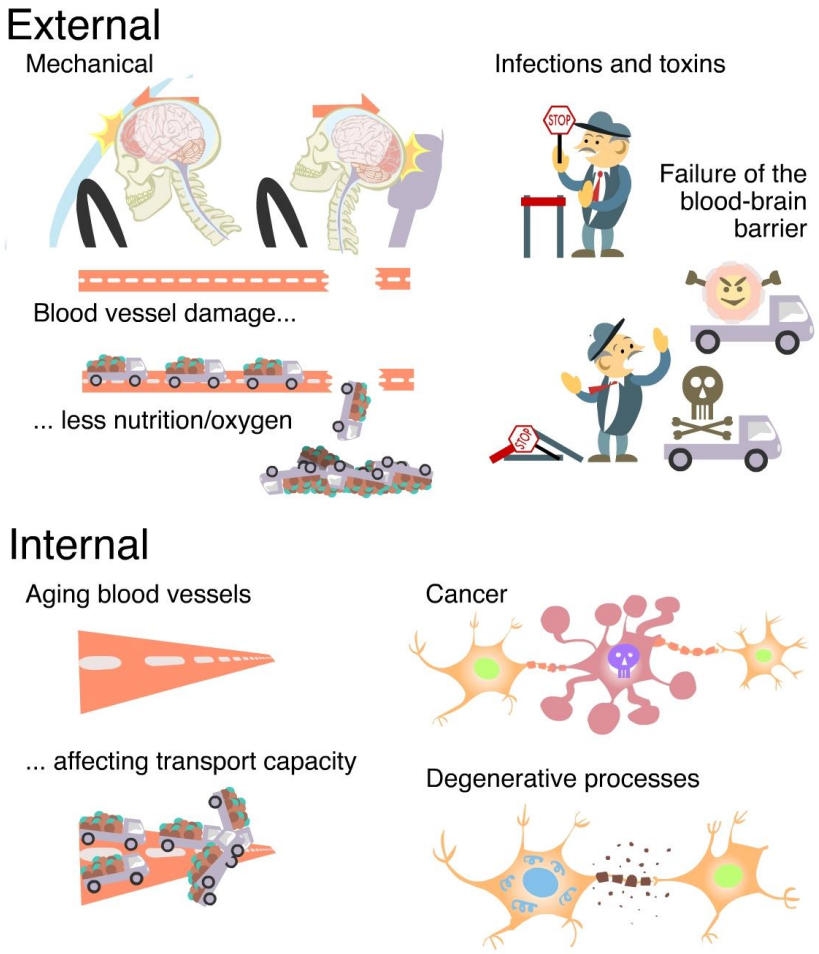
Which hospital department takes care of what depends on for example, the type of injury, the severity of the condition, the likelihood of survival (conditional of treatment), the age of the patient, etc. Figure 12 gives examples of causes. While symptoms in psychiatry, neurology and neurosurgery all emanate from the brain; distinctions between fields have varied over time.

The field of brain disorders used to be more unified, but is now often administratively divided between psychiatry and neurology. This division has been argued merited by imaging findings [47], yet also counter-argued [48], or even questioned given recent advances in neuroscience [49]. However, currently in Sweden, neurology primarily deals with disorders presenting somatic symptoms (e.g. multiple sclerosis, Parkinson's), psychiatry focuses on disorders of personality or affection (e.g. schizophrenia, depression), and memory clinics on patients with progressive, gradual and persistent (months, years) cognitive impairment.

Memory clinics have existed since the 1980s [50], following the growing notion that dementia was a disease rather than a special case of normal aging. Also, statistical classification of dementias has evolved, e.g. DSM-II mentioned senile and presenile dementia (but under “psychoses associated with organic brain syndromes”); the concept of “organic brain syndromes” lasted until DSM-III, when “dementia” was introduced [6]. Classification and administration can be seen to evolve parallel to refinements of methods of investigation.

Hypothetically, imagine making a differential diagnosis between psychosis and dementia via indirect observation (i.e. via presented symptoms only, prior to autopsy), without antipsychotics (developed in the 1950’s), or proper imaging of the brain (X-ray ca. 1900, but CT ca. 1970). This would have been hard, particularly in older patients, presenting more symptoms. Cases of younger patients with well-defined behavioral changes, where other diseases could be ruled out more readily, would have been easier (e.g. Rita Hayworth, Alois Alzheimer’s patient Auguste Deter, both in their early 50s [1, 6]. Still, better tools allow faster advance-

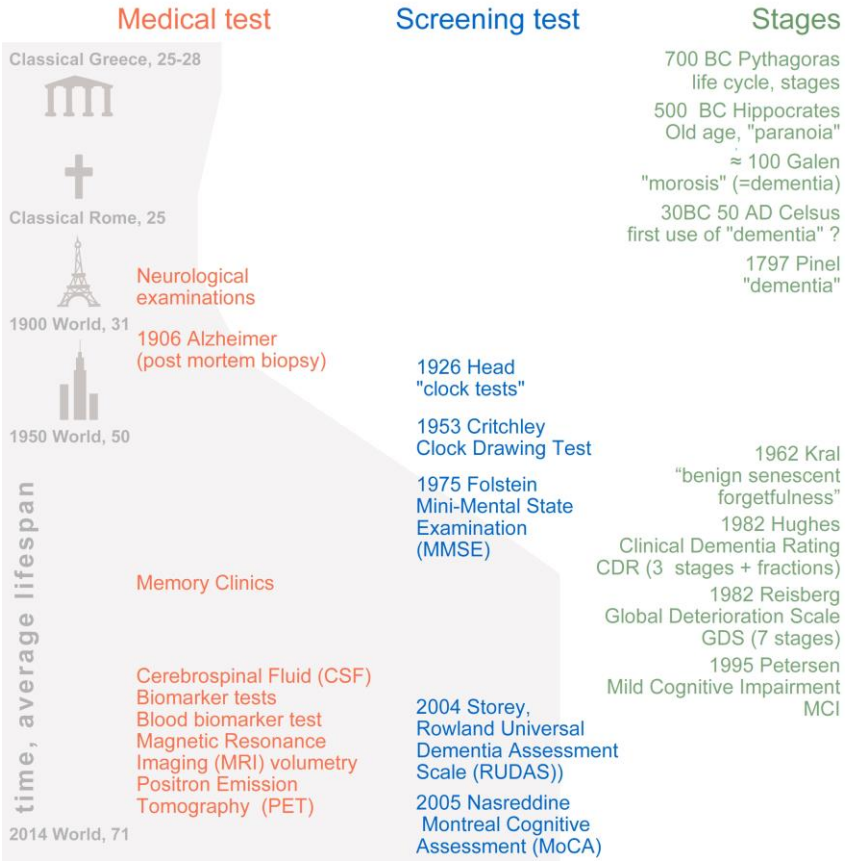
ment of science; and while some researchers e.g. expressed doubts about the concept of subcortical dementia in the late 1990's [6], the classification of subcortical ischemic vascular dementia was improved in 2002 [41].



12 Schematic causes of brain injury. ***External,*** *e.g. mechanical (wounds/shaking, e.g. causing shearing damage), infections/toxins passing the blood-brain barrier.* ***Internal,*** *may be vascular (e.g. aging blood vessels, diabetes), or cancer or degenerative diseases (e.g. Parkinson's, Alzheimer's).*

1.3.2 DEMENTIA

"Dementia" is not one disease, but a syndrome characterized by cognitive deficits that interfere with independence in activities of daily living, regardless of (organic) cause. Many brain injuries may lead to dementia (e.g. trauma, cancer, stroke). Age related cognitive decline has long been described, but finer nuances are of more recent date [6]. Figure 13 outlines a simplified timeline in relation to average lifespans.



13 Timeline of the history of dementia [6]. As the average lifespan increases so does the number of screening tests, dementia stages and classifications. Dementia goes from something "natural" to something pathological.

1.3.3 THE MEMORY CLINICS

Memory clinics are specialist clinics for patients who suffer from, or who appear to be at risk of developing, dementia. Common symptoms include loss of memory, loss of orientation, wordfinding problems, loss of ability to solve even simple emergencies. Diagnoses are often made in the earlier stages of dementia, not only to e.g. allow planning [50], but also as a smaller number of causes for dementia-like symptoms are partly reversible conditional on swift treatment (e.g. severe nutritional deficiencies, brain infections, subdural hematoma, normal pressure hydrocephalus [51], depression [52]). Figure 14 describes common examinations at a memory clinic.

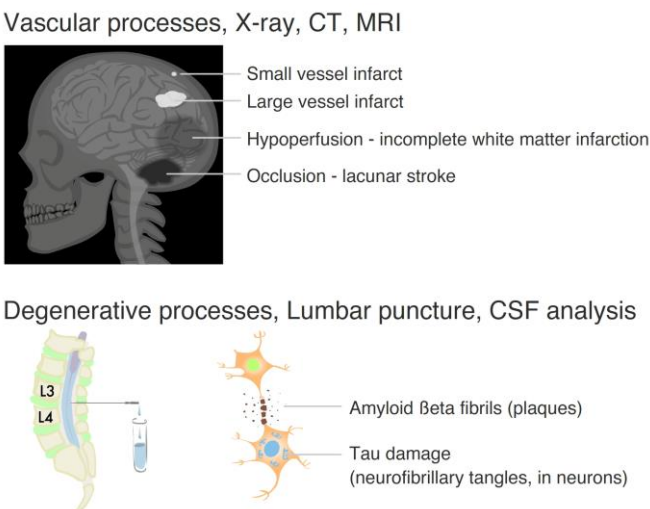
14 Common examinations in a memory clinic: 1. Anamnestic interview, 2. Blood samples, 3. Imaging (e.g. Magnetic Resonance Imaging (MRI), 4. Measurements of electrical activity (Electro Encephalogram, EEG), 5. Lumbar puncture (a needle is inserted to tap cerebrospinal fluid, CSF, in the lower back), 6. Neuropsychological tests.

1.3.4 MEMORY CLINIC DISEASES

The dementia syndrome may be etiologically categorized as degenerative (e.g. Alzheimer's, Huntington's, Parkinson's) or nondegenerative (e.g. posttraumatic, infectious, or toxic dementia)[1]. The most common cause

for dementia is Alzheimer's disease (AD, ≈ 65% of cases [1], > 95% sporadic [53]); the second most common dementia is vascular dementia (≈15% of cases) [54]. In addition to these, there are several other specific dementias (e.g. Frontotemporal Dementia, Lewy Body Dementia). The dominating dementia risk factor is age, and vascular or mixed dementias are more common after 80 years of age [54]. With rising life expectancy the number of dementia cases is expected to rise from just under 50 million today, to over 130 million by 2050 [55].

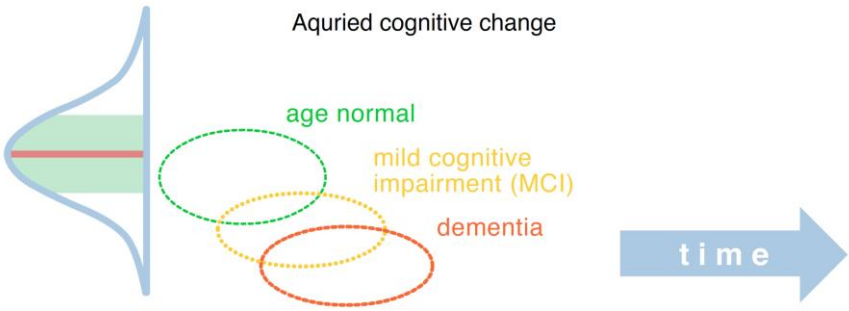
The exact "cause for AD" is currently not known, even if several theories exist (e.g. the amyloid cascade hypothesis [56]). Vascular dementias may emanate from several vascular related pathological actions [41], ranging from partial (hypoperfusion) to complete loss of vascular function (e.g. stroke). Common risk factors for vascular dementias are age, high blood pressure, diabetes, and a number of conditions affecting the cardiovascular system. The distinction between the dementias is further complicated by an increased risk for AD following years of vascular disease [41]. Patients in severe stages of dementia may appear more similar than patients in very early stages [42], and a severity staging assessment is a common part of clinical assessments, tracking progression.



15 Visual and chemical, two common clinical sources of information. Modern imaging techniques may combine the two, suggesting where in the brain different chemical compounds are found.

1.3.5 CLINICAL CHANGE, STAGES OF DETERIORATION

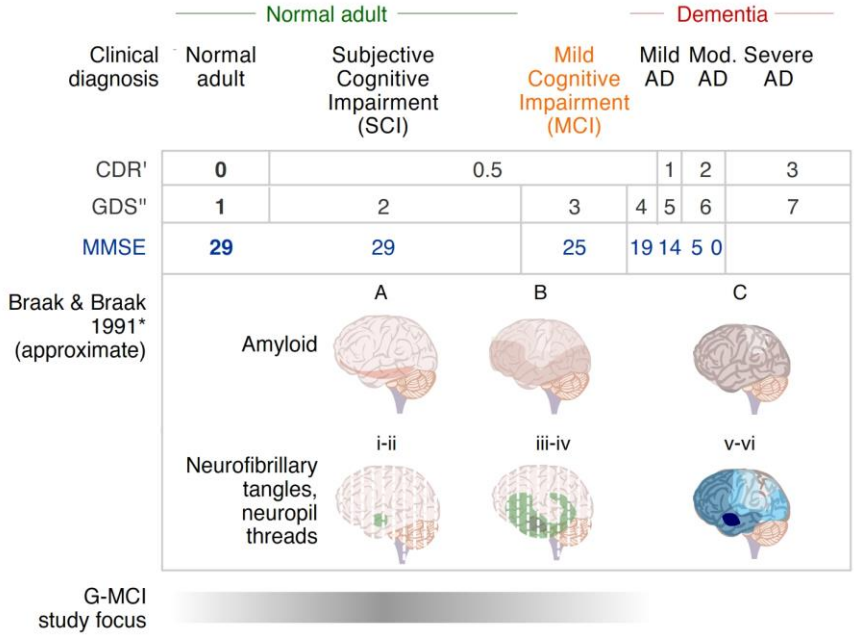
Dementia entails acquired (as opposed to developmental [6]) loss of cognition, and as normal cognition also declines with age, any pathological decline must present itself at a noticeably faster rate of decline. The intermediate zone of cognitively more declined than is normal for that age, but not yet demented, is referred to as Mild Cognitive Impairment (MCI) [57].



16 As cognition changes over time, so do the cutoffs for what may be considered pathological. Stages are overlapping.

Several staging systems exist, involving information from both patient and others, for example the Clinical Dementia Rating (CDR [58]) and the Global Deterioration Scale (GDS [59]) The CDR features a sum of boxes score from 6 different areas in a specific questionnaire of examination (memory, orientation, judgment & problem solving, community affairs, home & hobbies, and personal care). The GDS requires no formalized questionnaire, but also incorporates information from many sources.

For both scales, early stages may be superficially indistinguishable from age-normal functioning, but in the early-middle stages activities of daily life (ADL) start to fail and brain-imaging findings are common, e.g. decreased temporal areas and hippocampi [60], and in severe dementia the brain can no longer control the body. Figure 17 attempts to compare CDR and GDS with common findings on the Mini Mental Test, and brain changes described by Braak et al. [61].

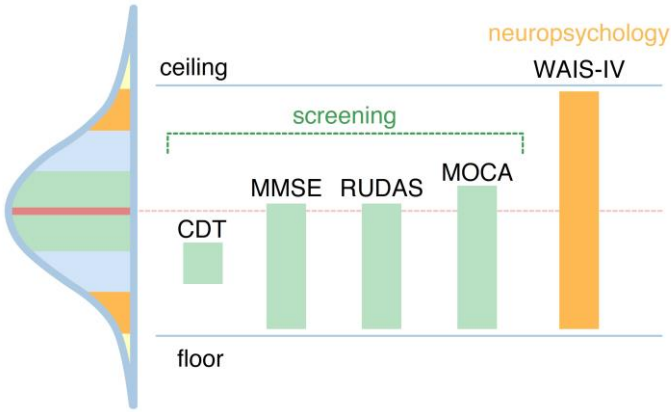


*17 Memory clinics may describe functional level in stages. The Gothenburg Mild Cognitive Impairment study (G-MCI) focuses on early stages [62]. At the later stages of dementia, the brain appears noticeably smaller with larger ventricles in both CT and MRI. Illustration adapted from Braak et al. 1991 *[61] and Reisberg et al. (2011) in [63]. '[58], ''[57]*

1.3.6 SCREENING TESTS, NEUROPSYCHOLOGY

Figure 18 suggests a few common cognitive tests' ceilings (the highest a test measures) and floors (the lowest a test measures). Screening tests are commonly used early in assessment. For impaired stages (from e.g. GDS 4, CDR 1, Figure 17), the screening tests Mini Mental State Examination (MMSE [64]) or the Rowland Universal Dementia Assessment Scale (RUDAS, [65]) suffice for classification, with the RUDAS less affected by language and possibly better for "ruling-in" dementia [65].

For milder stages the Montreal Cognitive Assessment Battery (MOCA [66]) may be used. For cases of early intervention, or cases where mainly subtle symptoms have been detected (e.g. SCI, MCI), neuropsychology adds information [67].



18 The highest a test measures is the ceiling, the lowest the floor. Screening tests' ranges approximated here are the Clock Drawing Test (CDT), the Mini Mental State Examination (MMSE), the Rowland Universal Dementia Assessment Scale (RUDAS), the Montreal Cognitive Assessment Battery (MOCA) and the Wechsler Adult Intelligence Scale (WAIS) IV.

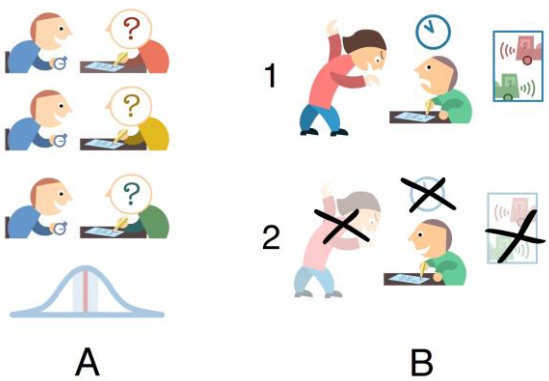
1.4 NEUROPSYCHOLOGY

A psychological test may be said to be a "sample of behavior obtained under controlled conditions" (p. 4, [2]) A cognitive test focuses on processes of cognition, and a neuropsychological (NP) assessment aims to paint a full and nuanced picture of a patient's cognitive capacity through weighing together several factors: the anamnestic history (including e.g. assessment of premorbid capacity, i.e. school grades, work history); possible psychiatric or other medical history; observations from interview and testing (including observations from e.g. spouse, friend, and reactions to stress from within the testing); scores from the NP tests compared to relevant normative material (as per the patient's age, level of education, etc.). Naturally, factors such as well-understood instructions and good motivation are of the utmost importance to the validity of the findings. It falls on the neuropsychologist to balance the normalized scores with the entire context they have been produced in. In doing so, the neuropsychologist may proceed in mainly one of two ways: a strictly

quantitative (actuarial), or a more hypothesis-testing, process-oriented fashion, also choosing between relatively fixed or flexible sets of tests [2].

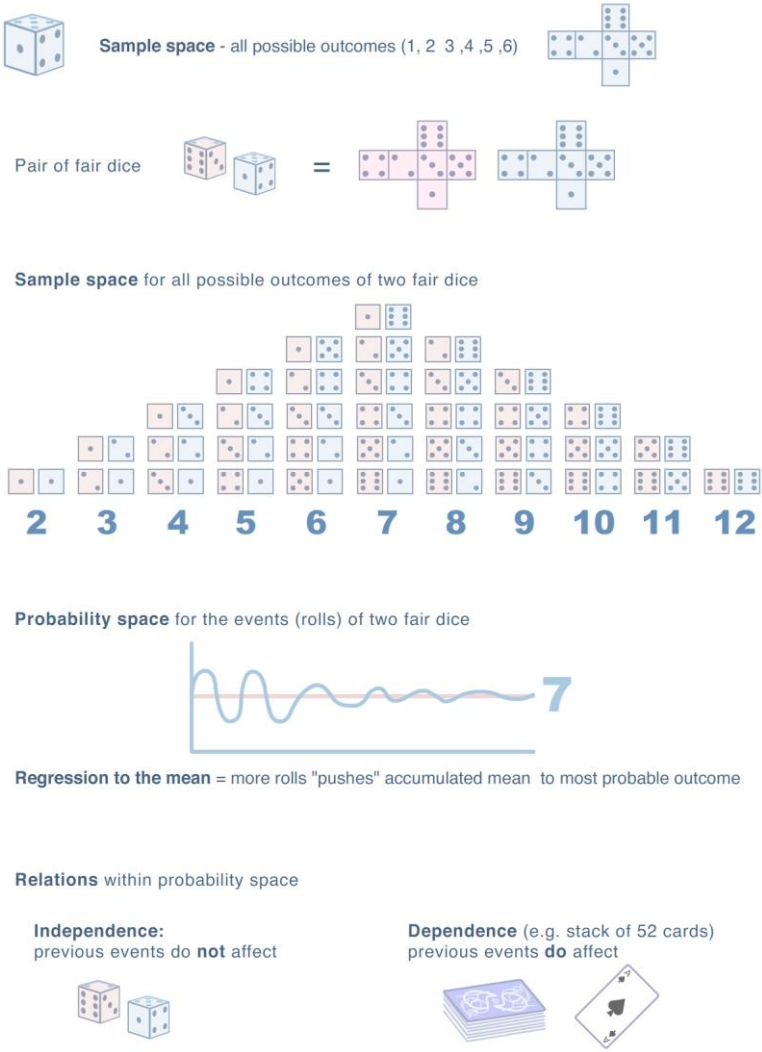
1.4.1 APPROACHES: ACTUARIAL VS. HYPOTHESIS-TESTING

An actuarial test administration (e.g. using exact instructional wordings) has benefits such as enabling identical repetitions, and less dependence on one particular neuropsychologist. A hypothesis-testing, process-based, administration, e.g. stepwise permitting use of tools, such as pen and paper in repeated administrations, conditional of failures, (aka "testing the limits" [2]), offers a deeper analysis and may better separate e.g. reported memory problems from attentional factors. While ideally an NP test should obtain the patients "best" possible performance, failures (e.g. shifting errors, sequencing errors [68, 69]), may be more informative to an investigation of disease, and different test approaches have different possibilities [70]. Hypothesis-testing may give deeper knowledge, but is less repeatable, relies more on the neuropsychologist, and will consume more time. Actuarial administration of a well-designed test (including its instructions) ideally relies less on any particular test administrator and may consume less time. Yet, while clinical neuropsychologists use both approaches, the unit of normalized measurement in classical test theory, the mean and standard deviation, is based on probability concepts [29].



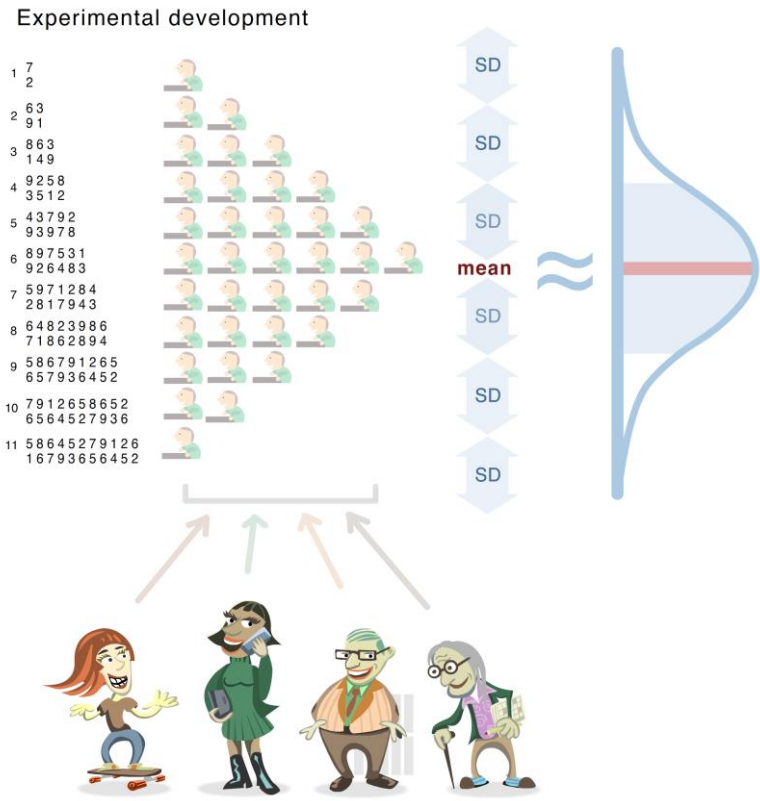
19 Actuarial (A) vs. hypothesis-testing (B). Identical administrations depend less on the test administrator. Hypothesis-testing may stepwise remove distractors and learn more of possible reasons for failures. Hypothesis-testing may in this experimental approach be said to be more in-between indirect and direct observation.

1.4.2 BASIC PROBABILITY CONCEPTS



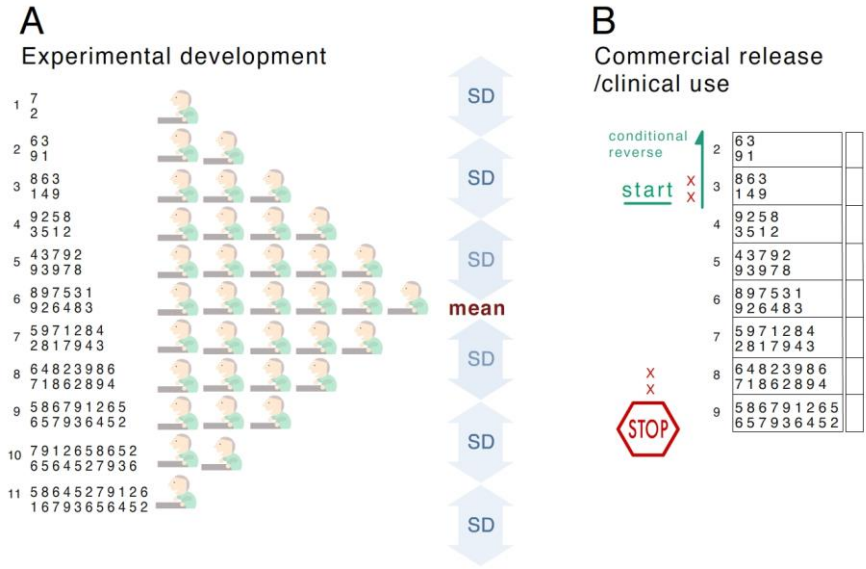
20 Basic probability concepts. A distribution of scores may be more or less similar to ideal distributions (e.g. those of a pair of fair dice). Statistical tests re-scale and compare observed distributions (scores from real administrations) to ideal distributions, for example the Student's T distribution (a distribution that varies as a function of "n", becoming similar to a normal distribution at around n = 30 and above). Many distributions are "normal" (Gaussian), but not all, see e.g. Micceri [71].

1.4.3 TEST DESIGN AND USAGE



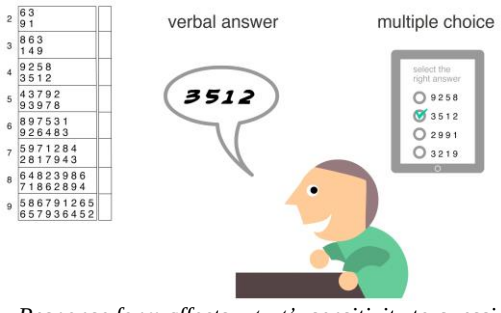
21 Schematic NP test development, part one. A task with good psychometric features is administered to a representative group of volunteers. The mean and standard deviation (SD) of the normative groups' scores are calculated. The standard deviation becomes the unit of measurement answering the question: "How far from the mean is one particular score."

Neuropsychological tests are developed through administration of a prototype test to several normative samples, e.g. participants of different ages and/or educational levels. When a cognitive task (e.g. digit repetition) has shown good psychometric features (e.g. reproducibility: reliability, and bearing to everyday tasks: validity) the test is adapted for clinical or commercial use. While normative administrations administer all items of a test design, commercial editions often aim to offer time saving devices.



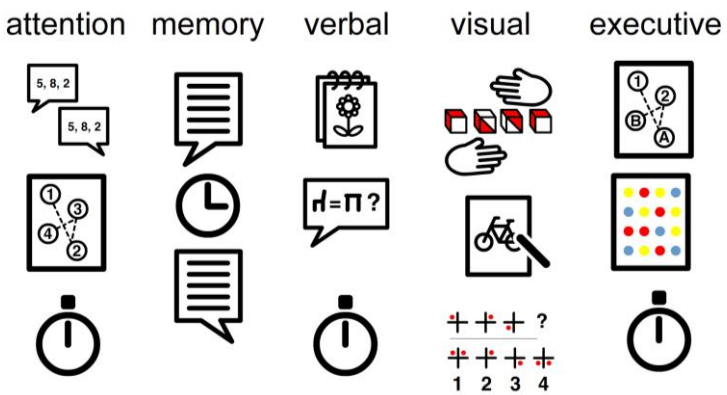
22 *Part two of the schematic development of NP tests, adding time saving features. The goal is to allow a shorter test administration (B) while retaining the ceiling-floor range of the original, experimental test (A).*

As individual test items may be arranged from high to low correct-answer-frequency, a starting point a few items into a test may be introduced with limited risk for erroneous classification, especially if combined with rules for reverse administration, conditional of errors. This way, a commercial test may retain a low floor and high ceiling while offering a shorter administration than in the development phase. Naturally, too many features from the initial test development cannot be changed without altering the possibility of obtaining scores by chance. For example, changing the way to respond from oral answers to multiple-choice may increase the chance of obtaining higher scores by guessing.



23 *Response form affects a test's sensitivity to guessing.*

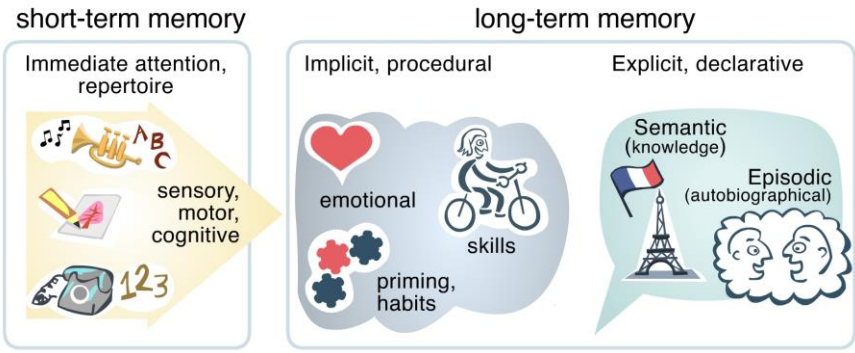
Currently, neuropsychology is more often used in assessing higher-level functioning (requiring interaction of many brain regions, e.g. Wechsler Logical Memory [29]). Some test batteries (e.g. the Halstead-Reitan) address more lower-level functions, include detailed examinations of left side, right side stimulation-response (e.g. Grooved Pegboard [29]), and are still used in special cases, for example in epilepsy surgery. However, with improved imaging techniques the need to psychometrically describe organic localization of injury has decreased [29, 72].



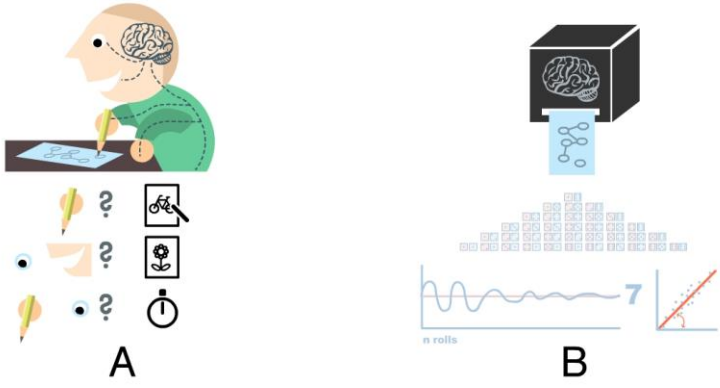
*24 Five common NP domains. A test is often sorted into a domain from features it **mainly** tests (domains may tap the same function [2]).*

1.4.4 COGNITIVE DOMAINS, THE DOMAIN OF MEMORY

For readability, neuropsychological reports often feature results structured in cognitive domains. Domains are not mutually exclusive (e.g. if verbal instructions are a part of a test, language and working memory will be). As mentioned above, cognitive domains do not present a one-to-one relationship to functional brain regions. While not practical to discuss in reports, neuropsychological domains are perhaps best thought of as approximate constructs, where a certain aspect may be argued in the report. The particular domains and tests of this thesis are further described and commented upon in the Materials and methods section, and in the Discussion. Further aspects of the domain of memory are shown in Figure 25.



25 Types of memory. Broadly, short-term/working memories are thought to consolidate into long-term [1, 73-75]. Episodic memory differs from semantic by including place and timestamp. As the hippocampi order memories via place cells, "time" becomes a special case of "place" (c.f. mental timeline).



26 Hypothesis-testing (A) may triangulate the source of a failure. A black box approach (B) attempts something similar through statistical means, by necessity more often used in research. "Confounders" may be investigated in the clinic, but are more often "controlled for" in research.

1.4.5 CONFOUNDERS – DISTORTING OR INFORMATIVE?

In research, result-affecting factors are sometimes referred to as "confounders" and by necessity "controlled for". In the clinic, a "confounder" may be produced from many things, including the test situation (e.g. "white coat hypertension", higher blood pressure only in the doctor's office [76]), and clinicians should aim to investigate if this appears to be

the case. In clinical situations "confounders" may be considered on a continuum: from pure distortion to informative:

- **Purely distorting factors**: Scoring errors; misunderstanding of test instructions (e.g. perceptual as well as language reasons); poor test design (e.g. overly aggressive termination rules); malingering (deliberately faking low scores).

- **Possibly informative offset scores**: Low scores produced from environmental factors. Shift work was reported causing not only sleepiness but also longer response times, and more errors (measured by an n-back and Continuous Performance Test) [77]. The time of day of NP testing may be of importance [29].

- **Somewhat distorting factors**: Transfer effects from previous experience, e.g. draughtsmanship from experience as an artist, [2], or from other test-similar tasks. Incidentally, the original Kohs block design was based on commercially available toys [10, 78], and it was later found that children who had played with the block design game "Trac 4" obtained higher scores, as did children who were allowed to play before testing [79]. Also, the structure provided by a test situation may affect the validity of the findings outside the examination room [2].

- **Practice effects**: Practice effects (score gains from repeated test-administrations) include all from direct learning to increased familiarity with the situation [80]. Practice effects are larger with shorter test-retest intervals, larger with performance tasks, and larger with younger persons, than with wider test-retest intervals, older persons and verbal tasks (Kaufman 1994 in [81]). Practice effects have been found larger at the first retest, compared to later [82]. A diagnostic value has been documented e.g. "practice effects on episodic memory tests were associated with a decreased risk of progression to AD" (abstract) [83] but also a confounding effect in that practice effects masked true decline [84]. Practice effects in motor control reduced movement jerk more in AD and MCI patients compared to controls [85]. However, practice effects may incorporate *both* a novelty effect (NE, to be "thrown" by a new task) and a learning effect [86]. And, as larger NE ("false baseline lows") contributed to larger practice effects,

separating NE from inability to learn may improve assessments [86].

- **Self-awareness of cognitive performance**: Cognitive capacity has been found to relate to awareness of said capacity [87]. Self-awareness has also been found to relate to pre-existing beliefs of cognitive capacity [88]. Self-report has been found to correlate moderately to test scores [70]. Taking a working memory test made participants feel older [89]. Subjective cognitive impairment (SCI), perceived cognitive loss and non-pathological NP test scores, has been found to correlate with stress [90]. "Diagnosis threat" may affect performance, as reminding patients of their neurological history was found to diminish subsequent NP performance [91], similar to effects from fear of AD [92]. Denial of problems may be a part in more advanced stages of MCI [93], or at least variation of self-awareness [94]. Comparisons between separately interviewed spouses and patients showed a sharp difference in complaints, beginning in mild dementia with spouses complaining more and patients less (Reisberg et al. 1985 quoted in)[95].

- **Person-to-person effects**: Hard for the individual neuropsychologist to explore, but a wide range exist: from perceived bias lowering scores [96]; to effects from "stereotype threats" (being at risk of confirming a negative preconception of one's group) impairing scores [97]; to the administrator showing subtle emotion (e.g. saying "fine" or nodding) increasing scores [98]; to changes in answering techniques, e.g. repeating the word list in RAVLT producing one word more [99]. Another category may be a neuropsychologist "slowly and unwittingly" [2] developing a certain administration style (e.g. slowly changing instructions) and blindness to this, a.k.a. examiner drift. If person-to-person effects follow the pattern of increased-anxiety in the examination [76] they may be attenuated by building trust, improving patient-examiner rapport. Depending on the clinical question, some of the above may be probed for further information or be regarded as distortion.

- **Second-language effects:** insufficient language skills will completely invalidate NP tests; modifications will invalidate norms

but may still produce valid inferences [29]. Interpreter use was found to increase verbal scores (Vocabulary, Similarities) the most and performance tests (Block Design, Matrix Reasoning) the least [100]. Bilingualism has been suggested to contribute to cognitive reserve [101], yet a publication bias favorable to positive findings has also been proposed [102].

Much of the added value of neuropsychology comes from considering the patient's entire context, not only the test scores (these may actually be a smaller part). The possibility to analyze errors, or e.g. repeat a test with and without a "confounder", is the clinics' largest advantage compared to research, particularly with regard to ecological validity.

1.5 INTRODUCTION KEY POINTS

- The history of cognitive measurement is relatively short.
- Cognition measurement is indirect observation.
- Brain network structure changes with age.
- A similar injury may have different effects depending on age.
- Comparable groups are used for normative assessment.
- Normative data are based on probability concepts.
- Tests have floors and ceilings.
- Staging systems suggest pathological deterioration.
- NP assessment balances between actuarial and hypothesis-testing.
- Neuropsychological domains overlap.
- "Confounders" range from distortive to informative.

1.6 KNOWLEDGE GAPS

1.6.1 TEST SECRECY AND MEMORY TESTS

As previous knowledge of a neuropsychological test might invalidate the results, neuropsychologists emphasize the importance of test secrecy. However, will test secrecy protect from memory training effects?

1.6.2 DO FREE CREDITS DAMAGE PRECISION?

As outlined in section 1.3.3, ideally a good test design retains a development version's ceiling-floor range via start-and-reverse, and termination rules in combination with free credits for items not administered. However, for the Boston Naming Test (BNT), do free credits affect scores identically for all stages of impairment?

1.6.3 PRACTICE EFFECTS: SIGNAL OR NOISE?

As memory clinics assess cognitive deterioration, what is the added value of NP follow-up and change scores (Δ-scores)? Do e.g. repeated test administration risk practice effects even in early stages of possible dementia?

1.6.4 NON-NATIVE SPEAKER: NATIVE NORMS OR NOT?

When assessing non-native speakers, what should guide use of native norms or not? Are second language effects mostly restricted to vocabulary tests? If verbal fluency assessed by a short conversation is not enough to merit use of native norms, how could a neuropsychologist proceed?

2 AIMS

The general aim was to investigate both why a direct measure of cognition is not possible, but also to what degree such an attempt could be informative. This was narrowed down to how four simplified assumptions may render neuropsychology less informative. The specific objectives became to investigate the following:

2.1 TEST SECRECY AND MEMORY TRAINING

Will test secrecy protect from memory training effects?

2.2 EFFECTS OF FREE CREDITS IN BNT

Will mixing free-credits and full-length BNT administrations matter?

2.3 WHAT WILL Δ-SCORES ADD?

Is noise from practice effects in repeated testing negligible? How do NP change scores (Δ-scores) differ between different clinical stages of cognitive decline and transitions between them?

2.4 SECOND LANGUAGE EFFECTS

What are the performance differences in native vs. non-native, Swedish speakers on a Swedish language administrated NP test battery?

3 PARTICIPANTS AND METHODS

3.1 PARTICIPANTS

3.1.1 PAPER I – THE CASE OF A MEMORY ATHLETE

Participant in paper I was one female 20-year old student trained in mnemonic techniques since the age of 12. The interview indicated no innate superior mnemonic capacity. The participant was contacted in connection with a public world record attempt and - while experienced in memory contests – had not been administered neuropsychological tests prior to the case study. Written informed consent was given to publish results of neuropsychological testing in anonymized form September 16, 2010.

3.1.2 PAPERS II, III - PARTICIPANTS IN G-MCI

Participants in papers II and III were participants of the Gothenburg Mild Cognitive Impairment study (G-MCI). Paper II, the Boston Naming Test (BNT) analyses, included 23 controls and 259 patients, and required full (60-item) BNT administration. Paper III, the change scores (Δ-scores) analyses, required that participants hade been assessed two times and included 64 controls and 470 patients. G-MCI exclusion guidelines were inclusion age outside of 50-79, prior head trauma, substance abuse, current psychiatric ailment (e.g. severe depression), or (for patients) symptom duration shorter than 6 months. Controls should present no cognitive complaints; have an MMSE at or above 26 (plus the same exclusion criteria as patients). Patients were recruited at the Sahlgrenska University Hospital Memory Clinic. Controls volunteered at, for example, information meetings on dementia. The G-MCI study was approved by the regional ethics board of the University of Gothenburg, diary number L091, March 15, 1999.

3.1.3 PAPER IV, PARTICIPANTS IN SCAPIS.

Participants in paper IV were recruited from the Gothenburg pilot part of the Swedish Cartdiopulmonary Bioimaging study (SCAPIS-pilot). Two hundred and thirty-seven were native Swedish speakers, 85 were non-native Swedish speakers. The entire SCAPIS project recruits a demographically representative set of 30 000 men and women between 50 and 64 years of age. Prior to the main SCAPIS, a feasibility study (SCAPIS-pilot) was performed 2012 in Gothenburg, inviting 2243 participants, recruiting 1111, from which the above participants were invited. Exclusion criteria were pathological NP and CDT scores and/or testing in another language than Swedish. SCAPIS was approved by the ethics committee at Umeå University, and the additional cognitive tests were approved by the regional ethics board of the University of Gothenburg, diary number 734-13, October 10, 2013

3.2 THE NEUROPSYCHOLOGICAL EXAMINATION

3.2.1 PAPERS I, II, III (BASED ON G-MCI)

Paper I used English test versions. Papers II and III were performed in Swedish. Interview background material was used in Paper I, but not in Papers II and III. Papers I, II and III, addressed these domains (test order and comments in Table 1).

Non-divided attention/Speed: Parallel Serial Mental Operations (PASMO [44]) subtask: reciting the Swedish alphabet only; Stroop 1, naming colors of colored dots; Trail Making Test (TMT) A, draw a line between numbered circles; Wechsler Adult Intelligence Scale - Revised (WAIS-R) Digit Span Forward, repeating numbers read aloud, span length.

Executive attention: PASMO, following a recital of the Swedish alphabet (28 letters A-Ö) recite letters-numbers A-1, B-2 etc., throughout the Swedish alphabet [44] (similar to oral TMT B [29] but longer). Rey Complex Figure (RCF), copy time of a complex figure; Stroop 2, naming print color of printed words, time; Stroop 3, naming colors of color words, time; TMT-B, draw a line between circles with alternating numbers-digits, 1-A-2-B etc., time; WAIS-R Digit Span Reverse, repeating numbers backwards, span length; WAIS-R Digit Span, total points of

forward and reverse; WAIS-R Symbol Digit coding, pencil symbols in empty spaces guided by numbers, points.

Learning-Memory: Rey Auditory Verbal Learning Test (RAVLT) immediate and delayed recall of 15 words, total sum of 5 learning trials (5*15 words), recognition (custom: 15 lines of 3 words with 1 target and 2 phonetically alliterative distractors); RCF immediate recall, delayed recall of the previously copied figure; Wechsler Logical Memory (WLM) immediate recall, delayed recall of two short stories.

Visuospatial: RCF copy, figure copy total points, copy strategy (A, full perception of the whole figure = 3p; B, partial perception = 2p; C, erratic perception and copy = 1p); VOSP silhouettes, recognition of skewed silhouettes; WAIS-R block design, recreate a pattern with two-colored plastic cubes, points with original speed bonuses. Draw a bike (Paper I only).

Verbal: Boston Naming Test (BNT), naming of pictures, Paper I incorporated free credits [29, 103] but Paper II analyzed several versions [104] and Paper III used only points from 30 (no item 50, 51 [104]); Controlled Oral Word Association Test (COWAT) verbal fluency letters F-A-S (3 x 1 minute, total sum); Token Test (re-positioning of plastic tokens from verbal instruction, 22-item form); WAIS-R Similarities (explain similarities). For Paper I COWAT, FAS was administered in writing in German.

3.2.2 PAPER IV (ADDITIONS TO SCAPIS PILOT)

Simple speed/attention: TMT A; Stroop Test Victoria version part 1 (colors); RCF copy time.

Divided (executive) attention: Stroop Test Victoria version part 2 (color or words), part 3 (color of color words); TMT B; Symbol Digit Modalities Test (SDMT) CAB version [105] write numbers according to symbols, same symbols but different numbers c.f. the original [29]; PASMO.

Learning/Memory: Short story memory test with repetition, same text as in CAB [105] but revised administration, allowing both verbatim and synonym answers for both immediate and delayed recall. RCF immediate, delayed recall, and recognition.

Visuo-constructive: RCF figure copy.

Verbal: Token test CAB 6 item version [105], similar to Token test [29] but shorter, verbal instruction to re-position one of 8 plastic "tokens" in relation to the remaining 7; COWAT FAS; Category Fluency Test “Animal Naming”, naming as many animals as possible in one minute; BNT-CAB, 30-item naming task with images redrawn from the original BNT [105, 106].

Further references to the above tests may be found in respectively [27, 29, 37, 44, 62, 106, 107].

3.2.3 TEST ADMINISTRATORS

Neuropsychological tests were performed by licensed psychologists, psychologists in training, or other researchers, under supervision of licensed psychologists. All tests were administered in Swedish – except for paper I where tests were administered in English. No formal assessment of eyesight or hearing was performed.

3.2.4 INTERVIEW, COMMENT ON MEMORY

While all participant assessments started with interviews, these interviews mainly served to gather basic information, getting acquainted, settling in in the room, etc. While information on personal memories could surface, autobiographical memory was not formally analyzed.

The domain of memory only addressed working memory, before and after distraction and/or within a timespan of ca. 20 – 40 minutes (Discussion).

3.2.5 DESELECTION OF NEUROPSYCHOLOGICAL TESTS

The G-MCI has continually evaluated test prototypes and new translations for possible inclusion in the study. Thus, in papers I-III, prototypes, unpublished translations, or tests only administered to a minority of participants, were excluded. Also, as the G-MCI is a clinical study, a small number (< 10) of administrations were offered in a patients’ native language (Finnish, English), and these administrations were excluded.

Table 1. Order of administration in papers I, III, III

Test, comment	**Paper I**	**Test order, Papers II, III**
		Session 1
BNT**	1	1
RAVLT learning, first recall	2	2
PASMO (Paper I, only to "Z")	3	3
TMT A, B	4	4
Draw a bike (not analyzed in Papers II, III)	5	5
WAIS III/R*** Digit Span	6	6
WAIS-III/R*** Block Design	7	7
RAVLT recall and recognition	8	8
WAIS-III Similarities		9
	15 min break	Session 2
WMS R – WLM first recall (two short stories)	9	1
RCF/RCFT copy	10	2
COWAT FAS^	11	3
RCFT (first recall)	12	4
VOSP (subtest II)	13	5
WAIS-III Letter-Number	14	6
Stroop (Victoria, 24 item version)	15	7
RCFT (delayed recall + recognition)	16	8
WMS-R – WLM delayed recall (both stories)	17	9
WAIS-III/R*** Digit Symbol-Coding	18	10
WAIS-III Digit Symbol-Coding, Incidental Learning	19	_
Token Test, subtest V,	-	11
WAIS-III Matrix Reasoning	20	_
WMS-III Faces ^^	21	_
WAIS-III Picture Completion ^^	22	_
WAIS-III Picture Arrangement ^^	23	_
WAIS-III Object Assembly ^^	24	_

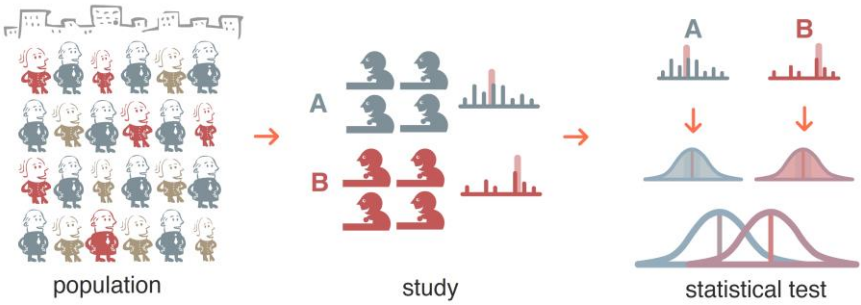
Abbreviations: *BNT, Boston Naming Test; RAVLT, Rey Auditory Verbal Learning Test; PASMO, Parallel Serial Mental Operations; TMT Trail Making Test; WAIS, Wechsler Adult Intelligence Test; R, Revised; WLM, Wechsler Logical Memory; COWAT, Controlled Oral Word Association Test; RCF/RCFT, Rey Complex Figure Test; VOSP, Visual Object and Space Perception Battery sub task 2 Silhouettes; WMS, Wechsler Memory Scale; ^ Paper I, Verbal Fluency (letters F-A-S, performed in native German, in writing). ^^ Paper I, Additional test Cf. G-MCI. ** Paper III, only BNT from 30 (no item 50, 51) *** Paper III, only WAIS-R. Please see text and Appendix for further details.*

3.3 STATISTICAL ANALYSIS, TESTS

Paper I did not feature statistical analyses per se, as it was a case study. Scores were compared to normative scores.

Paper II, III and IV used two-tailed Student's T-test to compare means of continuous variables, and Chi-square tests to compare dichotomous variables (comparing proportions to expected proportions). Bonferroni correction (a safeguard to retain a probability of 1-in-20 or 0.05 chance finding, by dividing 0.05 by the number of variables, in cases of many comparisons) was indicated where appropriate. In Paper III, NP change scores (Δ-scores, participants' raw follow-up scores minus raw baseline scores, per NP test) were compared to a hypothetical mean of 0. No imputation was used in any paper. For papers III and IV proportion of participants who completed a test were indicated as coverage percentage.

For paper III, a change algorithm was introduced, as the G-MCI GDS stages are ordinal but not equidistant. For any follow-up stage to also be classified as *changed* the following was required: significantly separated means and medians of MMSE and CDR-total: significantly separated mean and median total scores; *and* significant mean and median Δ-scores (mean Δ-scores ≠ 0, per Student's t test at $p < .05$, median Δ-scores ≠ 0 per Wilcoxon test).

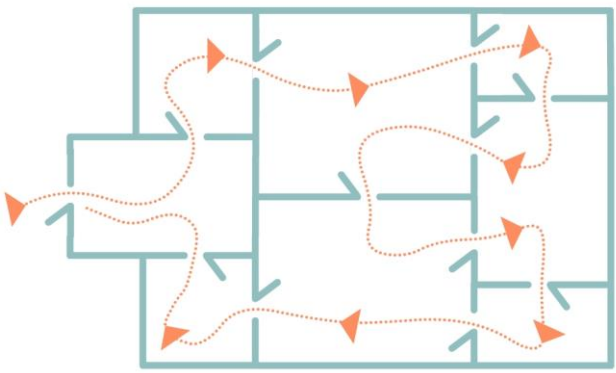


27 Basic principles of a statistical test (simplified: rescaling raw scores incorporating the number of participants). Overlap between groups A and B will affect the statistical analysis. Selection effects from the population will affect the validity of the findings.

4 RESULTS

4.1 PAPER I: MEMORY TRAINING OFFSET SCORES

Paper I aimed to investigate if test secrecy would protect from memory training effects. Paper I gives an example of how extensive memory training gave transfer effects that offset standard clinical tests of memory by 2-3 SD, in a young person who had trained for a long time. The memory athlete had trained the method of Loci [108] since the age of 12.

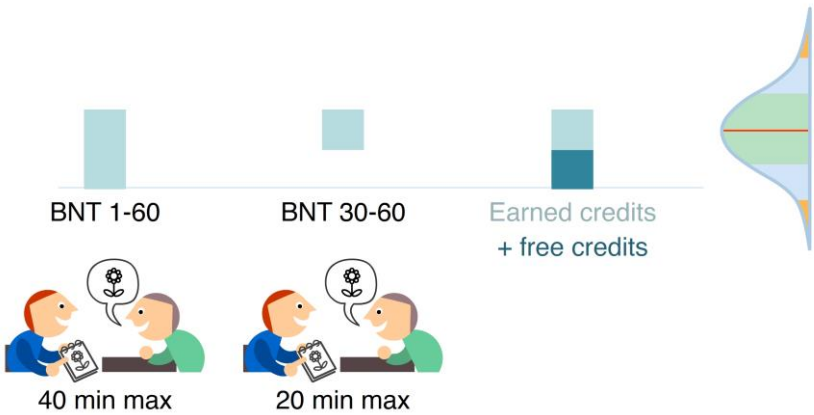


28 The method of Loci works through creating a "memory palace", a personal imagined physical location of high and various detail. Items to memorize are placed in different rooms and by "walking" an identical route every time, placing or picking up items, the capacity for fast storage of verbal information may be multiplied [108].

4.2 PAPER II: FREE CREDITS INFLATED SCORES

Paper II aimed to investigate if mixing free credits and full-length administrations of BNT would matter. Paper II illustrated how giving free credits for items not administered in the Boston Naming Test will most benefit those with the least chance of earning credits. Mildly demented patients' scores were found most inflated by free credits.

Note: a mix of scores (e.g. with later demented patients first administered the full BNT, but at follow-up administered the abbreviated version with free credits) was separately found to produce artifacts, i.e. erroneously suggesting that deteriorated patients had improved in vocabulary. The G-MCI database was subsequently amended to only allow use of the smallest common denominator: BNT administered from item 30 not counting items 50 and 51, as answers were changed after the publication of Tallberg's Swedish norms 2005 [109]). The above information was used in Paper III.



29 The Boston Naming contains 60 items, but may be administered from item 30 to save time. In those cases free credits are given for items not administered, so that the same set of norms may be used regardless of administration type. Paper II showed that such free credits inflated the scores of those most deteriorated.

4.3 PAPER III: PRACTICE EFFECTS WERE SMALL

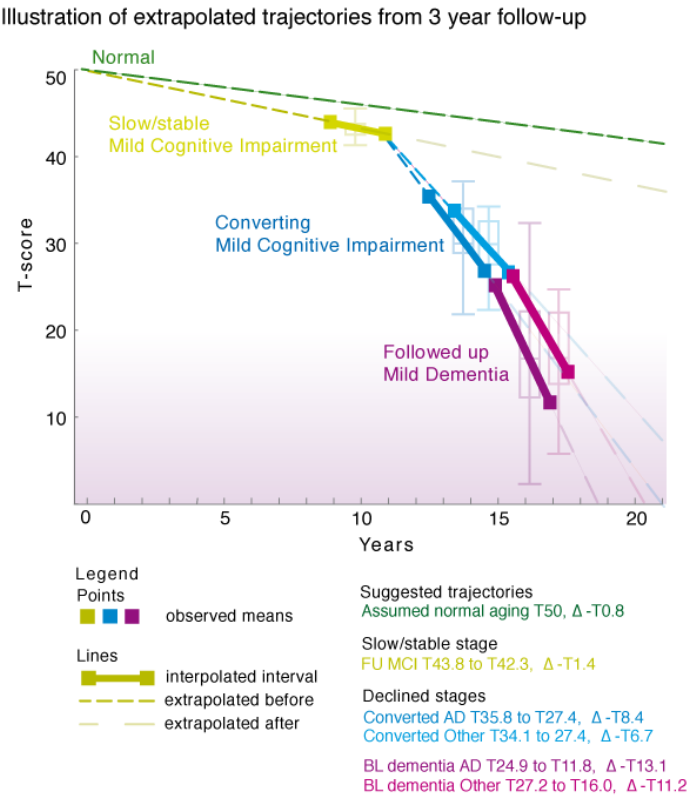
Paper III aimed to investigate if noise from practice effects in repeated testing would be negligible, and how Δ-scores differ between different clinical stages of cognitive decline and transitions between them. Paper III illustrated few significant discrete Δ-scores but also that (expectedly) fewer tests were administered, at both baseline and follow-up, to patients with more advanced deterioration (Figure 30).

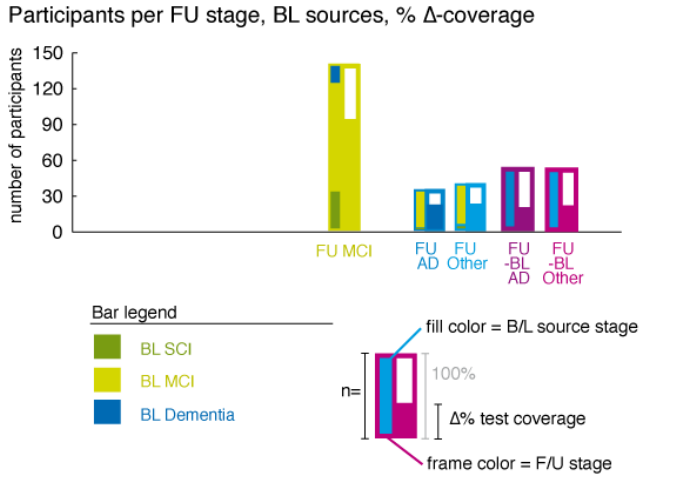
- Practice effects were not found large enough to recommend clinical decisions based on the "absence of practice effects".

- More impairment meant greater variability, thus mean Δ-scores described change better than Δ-scores for separate tests.

- Mean two-year Δ-scores in excess of 0.5 SD were only seen in patients converting to, or progressing in, dementia. A two-year cutoff of a 0.5 SD loss will likely work for memory clinic assessments.

- Practice effects could not be ruled out in MCI (Discussion).

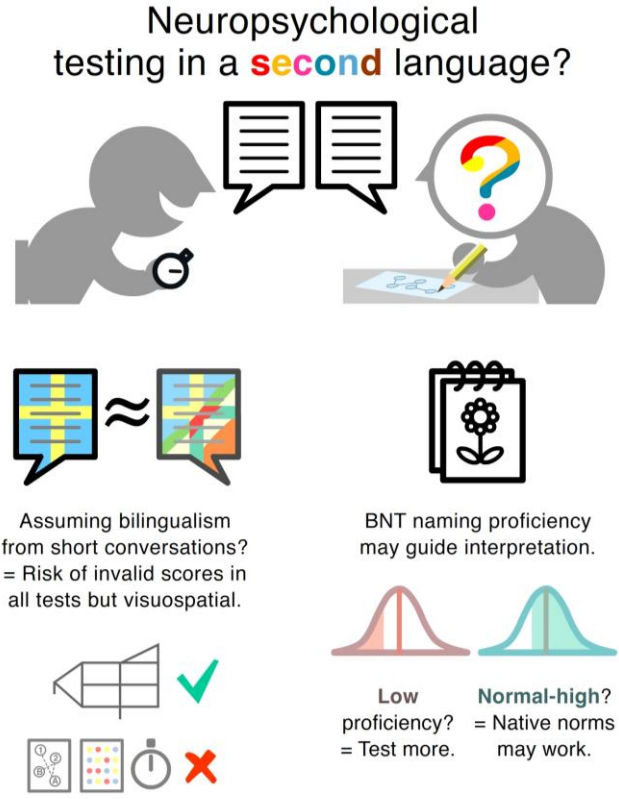




30 Abbreviations: FU, follow up; BL, baseline, AD, Alzheimer's disease; MCI, mild cognitive impairment. Top: extrapolated trajectories per stage with negative mean Δ-scores. Bottom: number of participants per stage, including baseline sources and Δ-test coverage (% identically repeated tests).

4.4 PAPER IV: 2:ND LANGUAGE HAD LARGE EFFECTS

Paper IV aimed to investigate if the performance differences in native vs. non-native Swedish speakers on a Swedish language administrated NP test battery. Comparing native and non-native Swedish speakers, we saw lower scores in many tests commonly thought to be tapping speed/attention. Assuming bilingualism from a short conversation was found inferior to assessing Swedish language proficiency via BNT (CAB, 30-item-version). For non-native speakers, younger age of arrival in Sweden, or arrival from a country with a language closer to Swedish (or where Swedish was also spoken) all contributed to higher NP-scores. Second-language effects were not found restricted to "verbal" tests.



31 Crib sheet for NP testing in a second language. Assuming bilingualism from short conversations is not recommended.

5 DISCUSSION

The specific aim of this thesis turned into investigating how simplified (but time-saving) assumptions may render neuropsychology less informative. Four assumptions were found to have potentially large effects: mnemonic training offset memory tests; free credits impaired BNT precision; Δ-scores were relatively noisy; second-language effects may be substantial.

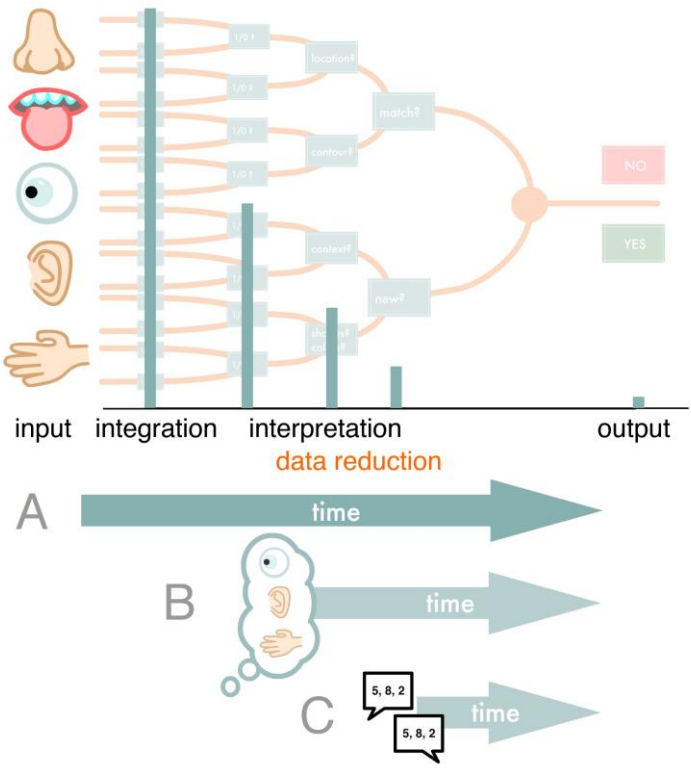
5.1.1 PRACTICE AND SPEED

Papers I and IV may appear dissimilar in focus, but they both address the effects of practice. Whereas most people do not train the method of loci for 8 years, they do invest more than 8 years when mastering a native language. In Paper I, training produced memory test ceiling effects, while in Paper IV a *lack* of training produced "slowness". However, importantly, non-native speakers in Paper IV neither reported ailments, occupational problems, nor differed from native speakers on visuo-constructive NP tasks. Further, inversely, the mnemonic master in Paper I did not report increased general memory capacity consistent with the exceptionally superior NP scores. A probable hypothesis is that both Paper I and IV give examples of partially invalid NP scores in terms of ecological validity, in cases of the usefulness (or not) of extensive training.

The ability for NP tests to predict everyday cognitive function in the outside world (ecological validity) has been found to be low to modest [70, 110], for some tests also relating to analysis of total scores or specific errors e.g. sequencing vs. shifting errors (e.g. TMT B [68, 69]). Many tests used in Papers I and IV were a) developed to detect disease, and/or b) involved test of "speed" (either e.g. TMT A, or as when story memory tests were read aloud at a certain pace). Clinically, a measure of "speed" is a rational choice as it creates a wide catch-all, and many factors may contribute to a "slow" result. Yet, for non-clinical situations, in findings of a "slow result", absence of evidence is not evidence of absence.

Figure 32 suggests how basically any overlearned mental skill (e.g. method of loci, native language) may serve as a shortcut, and contribute

to an impression of speed. And, inversely, several *lacks* of overlearned skills may give the impression of slowness; and multi-component tests may fail for many reasons [69]. In cases of dramatically different initial conditions (Papers I and IV), simple assumptions may result in lower ecological validity.



32 Operating from raw sensory input (A) will be slower than operating from integrated input (B) - but fastest will be overlearned skills (C). Increased modularity will save time and energy - but will also be more vulnerable. There are several forms of centrality in a multi-hierarchical system, and while injury to some areas (e.g. Thalamus, Frontal lobes) produces distinct symptoms this does not inversely prove the existence of one "central executive", as also argued in double dissociation: the observation that damage in one brain area causes certain deficits, does not rule out contributing problems from other areas (Teuber quoted in [72]).

5.1.2 FREE CREDITS IN BNT

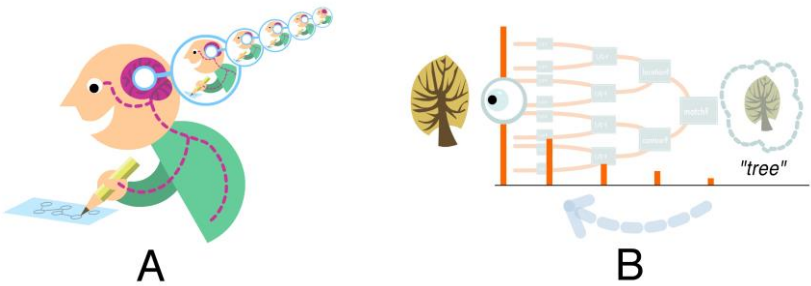
Paper II showed that free credits in BNT inflated the scores of demented patients. While free credits in combination with reverse and termination rules are not uncommon in NP (many WAIS tests have them [27, 37]), few NP tests offer as many free credits as BNT (30 of 60). Logically, for an omittable items + free credits system to work, the possibility of a person to earn credits must first be assessed. It might appear in cases of 100% reliable difficulty order, that omitting items in administrations to individuals with zero risk of failure on omitted items could never distort total scores. However, even cognitively intact persons may miss on "simpler" tasks, due to e.g. lapses of concentration and/or motivation and error-analysis may offer more information. For BNT, the type of error has been found informative, e.g., semantic errors [111], differences after phonemic cues [112], as well as response latencies [113]. Yet, even if errors were *never* informative, to hand out 50% free credits without risk, the item difficulty order would have to be 100% consistent. This is not the case for BNT.

For the 60 picture-items in BNT several *different* orders of difficulty have been found, e.g. one in Sweden [109], and others for African Americans and for Caucasians [114]. The Swedish publisher (Hogrefe) is aware of this, but not allowed to change the order [115]. Linguistic features of African American Vernacular English (AAVE) [116] vs. other English varieties are not directly applicable to Swedish conditions, but even to expect a difficulty order for 60 words to remain consistent over 30 years does not appear realistic from studies of vocabulary tests. Analyses of the Swedish Scholastic Aptitude Test (SweSAT/Högskoleprovet) showed considerable (110 of 151 words) changes in word understanding from 2000 to 2011 when examining 915 491 test takers, connected to e.g. changes in reading habits [117]. Granted, the SweSAT words were more abstract than BNT-pictures, but similar processes cannot be ruled out.

For BNT, time-saving through free credits does not appear to merit the risks of distorting scores, while more careful error analysis shows promise. In the case of mixing administrations (full administrations to some, abbreviated to some) systematic errors may be introduced. A shorter naming test, with better error analysis, administered in its entirety would be preferable.

5.1.3 EXECUTIVE, HIERARCHICAL

Paper IV could not confirm any "executive advantage". As described retrospectively by Baddeley (2012), the central executive (CE) was an organizing part in a 1974 working memory model suggested by Baddeley and Hitch, originally thought capable of independent attentional focus [118]. Similar executive control has been suggested to explain bilingual attentional control [101, 119, 120]. However, introducing a CE causes homunculus problems and infinite regress (i.e. does the CE have another CE, ad infinitum [121]). Baddeley (2012) later saw this problem and recommended using the CE/homunculus as a marker, not a solution [118].



33 A), homunculus ("little man") explanations lead to infinite regress ("who controls the little man's little man"). B), network models illustrate hierarchical function with a less paradoxical explanation.

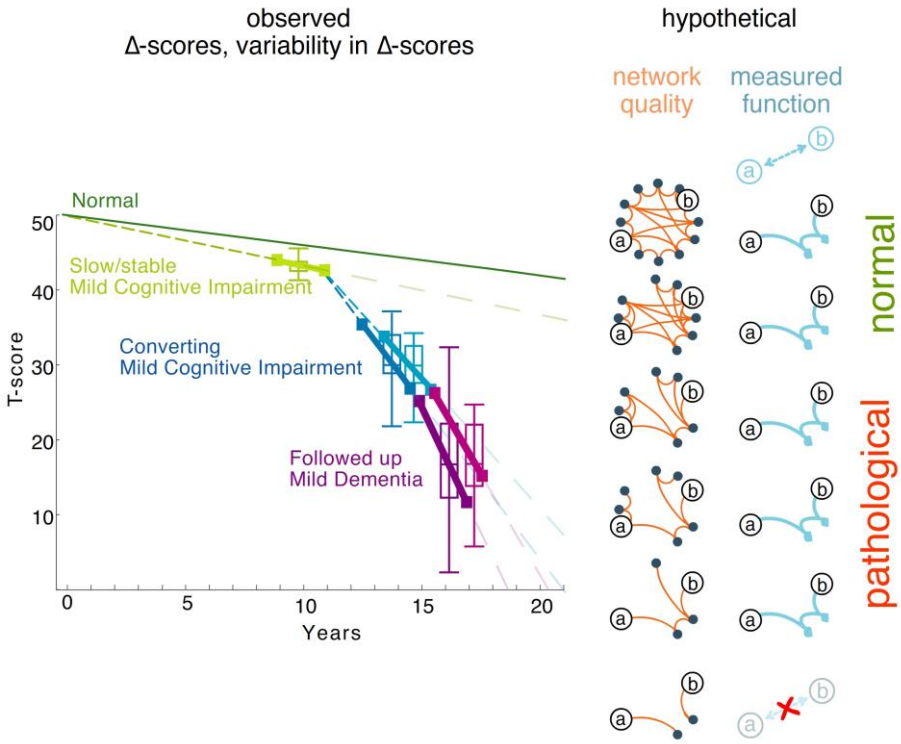
Further, findings have also been reported from many levels of language processing: listening in noisy environment was easier in L1 than L2 [122]; foreign languages sounded faster [123]; bilinguals made more tongue-twisting errors than monolinguals [124]; L2 vocabulary size was smaller [101, 125]; grammar was better in the language learned early [36]. Together, findings from different sources (motor, hearing, grammar) support a network model better than a homunculus theory, not to mention that different definitions of "executive" cannot even be compared [70]. We wanted to avoid the over-inclusive and potentially misleading term "executive function", and used "executive attention" in papers I, II, III, and "divided (executive) attention" in paper IV. We included the word "executive" for reader familiarity, the term is problematic, and if simple assumptions do not hold (as in Paper IV), more complex hypothetical constructs (i.e. "executive") will be even less valid. Presence of hierarchical principles does not infer a central executive.

5.1.4 "NORMAL" COGNITIVE AGING

For illustrative purposes, Paper III suggested a normal aging estimate that does not factor in cumulative capacities, such as learned languages. While there is discussion on the nature of cognitive reserve [126], there is agreement that speed and working memory change in a clearer fashion than verbal, habitual, or over-learned skills. A rough estimate from the domains of speed and learning suggests rounded up losses of around 2 SD over about 30-50 years [31, 32, 35]. Other sources suggest yearly losses of 2-3.5% of an SD for 50 years [127]. Yet, as variability increases with aging [34], as well as it manifests within-person in repeated testing [128], a generously rounded estimate is warranted. A rounded up estimate of 4 % of a SD of yearly loss would add up to 2 SD in 50 years, and likely cover most cases of normal aging, and was suggested in Paper III and Figure 34.

5.1.5 VARIABILITY: FROM GROUP TO INDIVIDUAL?

Paper III showed mean Δ-scores outside of controls' 2.5%- and 97.5%-iles were up to 10 times more common in declining patients. Specifically, the Δ-scores in paper III were produced by e.g. persons failing a task at one occasion, yet completing it at another, and this pattern was more pronounced in later stages of cognitive impairment. The link between group Δ-scores variability and observations of patients is the contrast between a skill remembered and a skill lost. While not specifically investigated (imaging data was not part of any paper), Figure 34 suggests how one function (a-to-b) may be present in all stages of deterioration as long as one connection remains, while the number of alternative functions decreases. A compounded effect will be an increasing gap between something that works (a-to-b) and capacities lost, similar to reported increase in variability in NP tests [34] and findings in Paper III. Further, hypothetically, an increased reliance on particular solutions could affect brain activation patterns, e.g. contribute to the reported increase in working memory activation in older persons [26]. Also, if a particular task (a-to-b) is overvalued and taken as a proxy for cognitive reserve, such "cognitive reserve" may confound: For example, a patient's use of a "difficult" word (a-to-b) may be more informative on discrete elements of spared function than of current general capacity. While potentially beneficial to patients, islands of "reserve" may confound NP assessments.



34 Findings in Paper III, contrasted to hypothetical network graphs. The observed means are group means, but similar variability was presented on a participant level (e.g. contrast between knowing one word but not another). Also, it may be argued that greater insult is not needed for greater injury: accumulated insult may give a critical mass effect.

5.2 LIMITATIONS

5.2.1 THE STREETLIGHT EFFECT

For all papers participant selection effects may have affected results, even if care was taken to analyze (e.g. compare the number of voiced concerns among native and non-native speakers, Paper IV) and/or amend this (e.g. exclude controls who developed dementia, Paper III). The NP tests for all papers were selected to cover commonly used domains, but the situation is still like Figure 35: after data collection we look under the streetlight. For example: for Paper I ecological validity was not formally tested. For Paper III autobiographical memories were not assessed; nor was perceived stress; nor motor control/learning, even though the latter

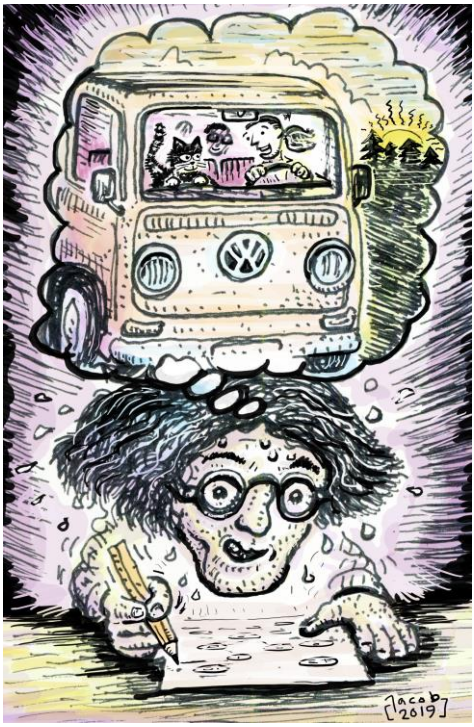
has been found informative [129, 130], both early with no clear memory deficits and later in disease progression [42], and with regard to practice effects [85]. For Paper IV results from test administration in native languages, and/or information of everyday functioning from e.g. informant reports/ questionnaires would have been valuable. Other aspects not investigated were e.g. eyesight and hearing, even if some aspects were implicitly noted (e.g. a patient repeating numbers correctly did hear them).

35 The streetlight effect. A policeman found a drunk looking for his lost keys and wallet under a streetlight. The policeman asked – "Why do you look under the streetlight?" – to which the drunk replied, "Because that's where the light is."

5.2.2 AUTOBIOGRAPHICAL MEMORY

To return to the initial patient-protest, "Yes, I saw you measure me, but deep down I know this to be impossible". The patient would be correct in many ways, for example: no *direct* measurement is possible. Also, for the neuropsychologists "memory" is testable memory, for a patient the word may refer to life events, autobiographical memory. No study in this thesis addressed autobiographical memory. Yet the nature of autobiographical memory also changes with age. Aging affects autobiographical memory with a "reminiscence bump" [131]: older adults (over 30-40 years) were found to more easily retrieve autobiographical memories from their 6-15 years, while young people more easily retrieved recent events [132]. Also, this telescoping effect seems to increase with age: interviews with 276 centenarians found that 70% remembered their most exciting event before the age of 40 [133].

Thus, between a young test administrator and an older patient, not only may cognitive capacities differ, but the entire perspective of life. Valid measures of reaction time will be informative to assess fitness to drive, but far from the complete measure of a person.



36 A hypothetical patient. Perhaps using the alphabet for the first time in 20 years, remembering what a driver's license once promised. Perhaps reflecting that most caregivers are half the patient's age.

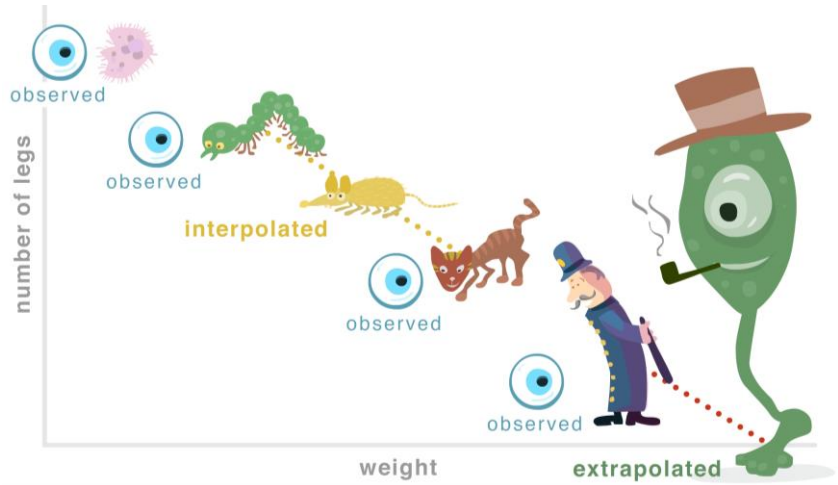
5.3 ETHICAL ISSUES

The NP Δ-scores illustrated in Paper III suggested different trajectories of decline for different clinical stages of memory clinic patients. However, even if this may appear to predict a remaining estimated useful life, health care is far from engineering. First, matters in engineering are interchangeable on another scale (e.g. iron is iron, but patients are unique). Second, NP scores may appear more precise than they actually are [29]. Third, the very act of NP measuring may have effects: e.g. cognitive tests affected healthy participants' feeling of subjective age (and positive

feedback did not help) [89], a moderate to high fear of AD impaired cognitive performance [92], and patients hearing about their own neurological history performed worse (diagnosis threat [91]).

Positive effects from optimistic expectations (placebo) have been documented in many forms (for example on pain [134]), but the opposite, negative effects from pessimistic expectations (nocebo) has only more recently been studied [135]. One way for nocebo to distort Δ-scores could be through initial stress at baseline producing "false lows" (stress was common in SCI [90]), with later follow-up perceived as less stressful, rendering seemingly improved scores (novelty effects [86]). As even effects from verbal suggestion have been found to have a measurable effect on cognition [135], nocebo effects cannot be ruled out. Possibly, such mechanisms could offer hypotheses as to why SCI has been found both a benign condition [136] and suggestive of further decline [137].

While dementia is very real and pathological NP scores document this precisely, to extrapolate beyond what has been observed is a delicate matter, especially if such guesses may aggravate symptoms or cause pain.



37 Schematic illustration of an assumed linear relationship between weight and number of legs. Extrapolation is guessing values outside the observed data points; interpolation is guessing values between them.

6 CONCLUSION

6.1 TEST SECRECY AND MEMORY TRAINING

Q: Will test secrecy protect from memory training effects?

A: Test secrecy did not protect memory tests from transfer effects from *extensive* training. World champions may be rare, but still.

6.2 EFFECTS OF FREE CREDITS IN BNT

Q: Will mixing free-credits and full-length BNT administrations matter?

A: Yes, free credits inflated Boston Naming Test scores of those most impaired. Mixing administration types will produce systematic errors.

6.3 WHAT WILL Δ-SCORES ADD?

Q: Is noise from practice effects in repeated testing negligible? Do Δ-scores differ between different clinical stages of cognitive decline and transitions between them?

A: Practice effects were too small to use the "absence of" for diagnostic purposes. But, only participants progressing to, or suffering from, dementia had mean Δ-losses in excess of 0.5 SD. For memory clinic use, a cut-off of a mean Δ-score loss of 0.5 SD per two-years may be sustainable.

6.4 SECOND LANGUAGE EFFECTS

Q: What are the performance differences in native vs. non-native Swedish speakers on a Swedish language administrated NP test battery?

A: That depends. If a non-native speaker's Swedish vocabulary has not been confirmed as normal-to-high (cf. Swedish native norms) results are more likely to feature invalid scores. Second-language effects were seen also in tests commonly thought to be tapping speed and attention.

7 FUTURE PERSPECTIVES

Neuropsychological testing will remain the gold standard to measure cognition. Yet, new technology offers many updates [138], and population changes (e.g. larger proportions of non-native speakers) and educational changes (e.g. less emphasis on handwriting, physical manipulation of objects, rote learning) will necessitate further study and development.

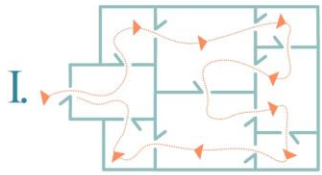
7.1.1 NEW DEVICES, NEW TESTS

Advances in computer technology introduce new inputs for neuropsychology. Computerized eye tracking may study what is visually focused on in relation to what is remembered. Automated speech analysis may offer higher resolution in timing of word fluency tests. Pad-administration of Trail Making Tests, recording the drawn line, may enable better error analysis. Computerized testing also opens for home examinations and more frequent follow up. Virtual reality may test capacities, e.g. for orientation, with greater ecological validity. Furthermore, computerized testing may be a better way to "save time" than free credits. Neuropsychologists can use the time saved to more carefully examine factors that are best manually assessed (e.g. motivation, fatigue, reasons for failures, types of errors, etc.).

7.1.2 NEW POPULATIONS, LANGUAGE LEARNING

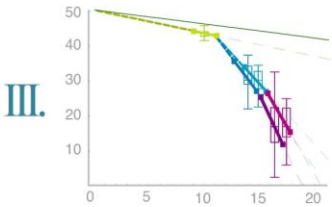
The number of speakers with any degree of multilingualism is now 50%, neuropsychology will need more non-verbal tests. Also, studies from bilingual countries (e.g. Canada) have long indicated that many factors affect how speakers of one language learn the other [139]. Future studies of Swedish second language effects should address socioeconomic class, attitude towards second language community, participation in culture of L1 and L2 communities, feeling of conflict between L1 and L2 communities, awareness of ridicule or shame from using one language etc. Participants in Paper IV had lived in Sweden for a mean of 34.8 years, and yet presented large differences in Swedish proficiency. Study of factors contributing to integration will be essential.

8 TAKE HOME MESSAGE

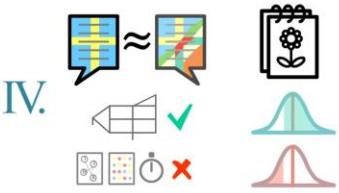


Memory training caused ceiling effects. Training effects may be large, even where not expected.

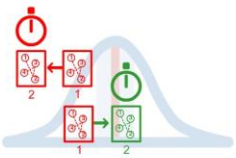
Free BNT credits did not add precision. Use a shorter test if pressed for time.



Variability increased in cognitive decline. Assessment via baseline scores and the mean of many change scores is better.



Test taking in a 2:nd language may affect all tests but visuo-constructive. Vocabulary assessment may guide use of native speaker norms.



More is more.
In clinical work, gathering more information will be more informative.
The added value of neuropsychology is the neuropsychologist.

38 *Take home messages per papers I-IV.*

8.1 IMPORTANCE

While increased accuracy is always valuable for research, the clinical importance of neuropsychological test precision lies in the benefits for the patient, not only for possible treatment and planning, but also for well being and useful support.

9 ACKNOWLEDGEMENTS

Many thanks to mum, dad (miss you dad) and sister for support and inspiration).

Thanks to my supervisors for all valuable advice: Anders Wallin, Per Hellström and Carl Eckerström.

Thanks to: Marie Eckerström, Niklas Klasson, Mattias Göthlin, Eva Bringman, Marie Johansson, Isabelle Rydén, Nora Balogh, Timothy Hadarsson, Helge Malmgren, Erik Joas, Sindre Rolstad, Idriz Zogaj, and the late Arto Nordlund.

Special thanks: Karin O´Toole Augustsson, Madielene Wetterskog and Madelene Axelsson.

Love to Selma the cat for being the kindest cat imaginable.

Finally,

Immense thanks to all volunteer participants who gave of their time and contributed to another piece of the cognition puzzle.

This work was funded by: Stiftelsen för Gamla Tjänarinnor, Stiftelsen Wilhelm och Martina Lundgrens Vetenskapsfond, Gun och Bertil Stohnes stiftelse, The Sahlgrenska University Hospital, The Swedish Medical Research Council, The Swedish Alzheimer Foundation, The Swedish Dementia Foundation, The Swedish Psychiatric Research Foundation, Hjalmar Svensson Foundation, Konung Gustaf V:s och Drottning Viktorias Frimurarstiftelse, The Gothenburg Foundation for Neurological Research.

Thank you very much!

10 REFERENCES

1. Kolb, B. and I.Q. Whishaw, *Fundamentals of human neuropsychology*. 5th ed. A series of books in psychology2003, New York, NY: Worth Publishers.
2. Vanderploeg, R.D., *Clinician's guide to neuropsychological assessment*. 2nd ed2000, Mahwah, N.J.: Lawrence Erlbaum Associates. x, 551 p.
3. Klein, H.A., *The Science of Measurement: A Historical Survey*2012: Dover Publications.
4. Nelson, R.A., *Foundations of the international system of units (SI).* The Physics Teacher, 1981. **19**(9): p. 596-613.
5. Smith, C.U., *Cardiocentric neurophysiology: the persistence of a delusion.* J Hist Neurosci, 2013. **22**(1): p. 6-13.
6. Boller, F. and M.M. Forbes, *History of dementia and dementia in history: an overview.* Journal of the Neurological Sciences, 1998. **158**(2): p. 125-33.
7. Sanchez, G.M. and A.L. Burridge, *Decision making in head injury management in the Edwin Smith Papyrus.* Neurosurgical Focus, 2007. **23**(1): p. E5.
8. Elhadi, A.M., et al., *The journey of discovering skull base anatomy in ancient Egypt and the special influence of Alexandria.* Neurosurgical Focus, 2012. **33**(2): p. E2.
9. Serageldin, I., *Ancient Alexandria and the dawn of medical science.* Glob Cardiol Sci Pract, 2013. **2013**(4): p. 395-404.
10. Boake, C., *From the Binet-Simon to the Wechsler-Bellevue: tracing the history of intelligence testing.* Journal of Clinical and Experimental Neuropsychology, 2002. **24**(3): p. 383-405.
11. Gregory, R.J., *Psychological Testing: History, Principles, and Applications*2004: Pearson/A and B.
12. Missios, S., *Hippocrates, Galen, and the uses of trepanation in the ancient classical world.* Neurosurgical Focus, 2007. **23**(1): p. E11.
13. Waal, F.B.M.d., *Are we smart enough to know how smart animals are?*2016, New York: W.W. Norton & Company. 340 pages.
14. Stiles, J. and T.L. Jernigan, *The basics of brain development.* Neuropsychology Review, 2010. **20**(4): p. 327-48.

15. Raichle, M.E. and D.A. Gusnard, *Appraising the brain's energy budget.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(16): p. 10237-9.
16. Futagi, Y., Y. Toribe, and Y. Suzuki, *The grasp reflex and moro reflex in infants: hierarchy of primitive reflex responses.* Int J Pediatr, 2012. **2012**: p. 191562.
17. Mengistu, H., et al., *The Evolutionary Origins of Hierarchy.* PLoS Computational Biology, 2016. **12**(6): p. e1004829.
18. Baum, G.L., et al., *Modular Segregation of Structural Brain Networks Supports the Development of Executive Function in Youth.* Current Biology, 2017. **27**(11): p. 1561-1572 e8.
19. Shore, R. and R. Shore, *Rethinking the brain : new insights into early development*1997, New York: Families and Work Institute. xviii, 92 p.
20. Ryan, T.J. and S.G. Grant, *The origin and evolution of synapses.* Nature Reviews. Neuroscience, 2009. **10**(10): p. 701-12.
21. Alberts, B., Dennis Bray, Karen Hopkin, Alexander Johnson, Julian Lewis, Martin C. Raff, Keith Roberts, and Peter Walter, *Essential Cell Biology*. 4th ed2014, New York: Garland Science.
22. Hoffman, P. and A.M. Morcom, *Age-related changes in the neural networks supporting semantic cognition: A meta-analysis of 47 functional neuroimaging studies.* Neuroscience and Biobehavioral Reviews, 2018. **84**: p. 134-150.
23. Reuter-Lorenz, P.A., et al., *Age differences in the frontal lateralization of verbal and spatial working memory revealed by PET.* Journal of Cognitive Neuroscience, 2000. **12**(1): p. 174-87.
24. Paige, L.E., et al., *Age differences in hippocampal activation during gist-based false recognition.* Neurobiology of Aging, 2016. **46**: p. 76-83.
25. Park, D.C. and P. Reuter-Lorenz, *The adaptive brain: aging and neurocognitive scaffolding.* Annual Review of Psychology, 2009. **60**: p. 173-96.
26. Cabeza, R., et al., *Maintenance, reserve and compensation: the cognitive neuroscience of healthy ageing.* Nature Reviews. Neuroscience, 2018. **19**(11): p. 701-710.
27. Wechsler, D. and A. Bartfai, *WAIS-RS : Wechsler Adult Intelligence Scale-revised : manual*, 1994, Psykologiförlaget: Hägersten. p. [2] s., s. 5-79.
28. Wechsler, D., *WMS-III svensk version*, in *Wechsler Memory Scale - third edition*, H. Nyman, Editor 2008, NCS Pearson Inc.: Stockholm, Sweden.
29. Strauss, E., E.M.S. Sherman, and O. Spreen, *A compendium of neuropsychological tests : administration, norms, and*

commentary. 3rd ed2006, Oxford ; New York: Oxford University Press. xvii, 1216 p.
30. Salthouse, T.A., *When does age-related cognitive decline begin?* Neurobiology of Aging, 2009. **30**(4): p. 507-14.
31. Hartshorne, J.K. and L.T. Germine, *When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span.* Psychological Science, 2015. **26**(4): p. 433-43.
32. Baxendale, S., *IQ and ability across the adult life span.* Applied Neuropsychology, 2011. **18**(3): p. 164-7.
33. Mitrushina, M.N., *Handbook of normative data for neuropsychological assessment*. 2nd ed2005, New York: Oxford University Press. xxii, 1029 p.
34. Wisdom, N.M., J. Mignogna, and R.L. Collins, *Variability in Wechsler Adult Intelligence Scale-IV subtest performance across age.* Archives of Clinical Neuropsychology, 2012. **27**(4): p. 389-97.
35. Park, D.C., et al., *Models of visuospatial and verbal memory across the adult life span.* Psychology and Aging, 2002. **17**(2): p. 299-320.
36. Hartshorne, J.K., J.B. Tenenbaum, and S. Pinker, *A critical period for second language acquisition: Evidence from 2/3 million English speakers.* Cognition, 2018. **177**: p. 263-277.
37. Wechsler, D., *Wechsler Adult Intelligence Scale - Third Edition.* Administration and Scoring Manual, 1997. **Third edition**.
38. Reuter-Lorenz, P.A. and D.C. Park, *How does it STAC up? Revisiting the scaffolding theory of aging and cognition.* Neuropsychology Review, 2014. **24**(3): p. 355-70.
39. Stern, Y., *What is cognitive reserve? Theory and research application of the reserve concept.* Journal of the International Neuropsychological Society, 2002. **8**(3): p. 448-60.
40. Rolstad, S., et al., *Biomarkers in relation to cognitive reserve in patients with mild cognitive impairment--proof of concept.* Dementia and Geriatric Cognitive Disorders, 2009. **27**(2): p. 194-200.
41. Roman, G.C., et al., *Subcortical ischaemic vascular dementia.* Lancet Neurology, 2002. **1**(7): p. 426-36.
42. Almkvist, O., *Cognitive syndrome(s) in preclinical and clinical vascular dementia.* International Psychogeriatrics, 2003. **15 Suppl 1**: p. 127-31.
43. Wallin, A., et al., *Alzheimer's disease--subcortical vascular disease spectrum in a hospital-based setting: Overview of results*

from the Gothenburg MCI and dementia studies. Journal of Cerebral Blood Flow and Metabolism, 2016. **36**(1): p. 95-113.
44. Nordlund, A., et al., *The Goteborg MCI study: mild cognitive impairment is a heterogeneous condition.* Journal of Neurology, Neurosurgery and Psychiatry, 2005. **76**(11): p. 1485-90.
45. Nagata, K., et al., *Clinical diagnosis of vascular dementia.* Journal of the Neurological Sciences, 2007. **257**(1-2): p. 44-8.
46. Gratton, C., et al., *Focal brain lesions to critical locations cause widespread disruption of the modular organization of the brain.* Journal of Cognitive Neuroscience, 2012. **24**(6): p. 1275-85.
47. Crossley, N.A., et al., *Neuroimaging distinction between neurological and psychiatric disorders.* British Journal of Psychiatry, 2015. **207**(5): p. 429-34.
48. Nair, A., *Neuroimaging distinction between neurological and psychiatric disorders--was there really one?* British Journal of Psychiatry, 2016. **208**(3): p. 298.
49. Baker, M.G., R. Kale, and M. Menken, *The wall between neurology and psychiatry.* BMJ: British Medical Journal, 2002. **324**(7352): p. 1468-9.
50. Jolley, D., S.M. Benbow, and M. Grizzell, *Memory clinics.* Postgraduate Medical Journal, 2006. **82**(965): p. 199-206.
51. Vakili, S., et al., *Timing of surgical treatment for idiopathic normal pressure hydrocephalus: association between treatment delay and reduced short-term benefit.* Neurosurgical Focus, 2016. **41**(3): p. E2.
52. Tripathi, M. and D. Vibha, *Reversible dementias.* Indian J Psychiatry, 2009. **51 Suppl 1**: p. S52-5.
53. Masters, C.L., et al., *Alzheimer's disease.* Nat Rev Dis Primers, 2015. **1**: p. 15056.
54. O'Brien, J.T. and A. Thomas, *Vascular dementia.* Lancet, 2015. **386**(10004): p. 1698-706.
55. Prince, M.J., et al., *World Alzheimer Report 2015 - The Global Impact of Dementia.* An analysis of prevalence, incidence, cost and trends, 2015.
56. Karran, E., M. Mercken, and B. De Strooper, *The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics.* Nature Reviews. Drug Discovery, 2011. **10**(9): p. 698-712.
57. Reisberg, B., et al., *Mild cognitive impairment (MCI): a historical perspective.* International Psychogeriatrics, 2008. **20**(1): p. 18-31.
58. Hughes, C.P., et al., *A new clinical scale for the staging of dementia.* British Journal of Psychiatry, 1982. **140**: p. 566-72.

59. Reisberg, B., et al., *The Global Deterioration Scale for assessment of primary degenerative dementia.* American Journal of Psychiatry, 1982. **139**(9): p. 1136-9.
60. Eckerstrom, C., et al., *Small baseline volume of left hippocampus is associated with subsequent conversion of MCI into dementia: the Goteborg MCI study.* Journal of the Neurological Sciences, 2008. **272**(1-2): p. 48-59.
61. Braak, H. and E. Braak, *Neuropathological stageing of Alzheimer-related changes.* Acta Neuropathol, 1991. **82**(4): p. 239-59.
62. Wallin, A., et al., *The Gothenburg MCI study: Design and distribution of Alzheimer's disease and subcortical vascular disease diagnoses from baseline to 6-year follow-up.* Journal of Cerebral Blood Flow and Metabolism, 2016. **36**(1): p. 114-31.
63. Abou-Saleh, M.T., C.L.E. Katona, and A. Kumar, *Principles and practice of geriatric psychiatry*. 3rd ed2011, Chichester, West Sussex England ; Hoboken, NJ: J. Wiley. xx, 910 p.
64. Folstein, M.F., S.E. Folstein, and P.R. McHugh, *"Mini-mental state". A practical method for grading the cognitive state of patients for the clinician.* Journal of Psychiatric Research, 1975. **12**(3): p. 189-98.
65. Rowland, J.T., et al., *The Rowland Universal Dementia Assessment Scale (RUDAS) and the Folstein MMSE in a multicultural cohort of elderly persons.* International Psychogeriatrics, 2006. **18**(1): p. 111-20.
66. Nasreddine, Z.S., et al., *The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment.* Journal of the American Geriatrics Society, 2005. **53**(4): p. 695-9.
67. Jansen, W.J., et al., *The Diagnostic and Prognostic Value of Neuropsychological Assessment in Memory Clinic Patients.* Journal of Alzheimer's Disease, 2017. **55**(2): p. 679-689.
68. Varjacic, A., et al., *Neural signatures of Trail Making Test performance: Evidence from lesion-mapping and neuroimaging studies.* Neuropsychologia, 2018. **115**: p. 78-87.
69. Stuss, D.T., et al., *The Trail Making Test: a study in focal lesion patients.* Psychological Assessment, 2001. **13**(2): p. 230-9.
70. Chaytor, N. and M. Schmitter-Edgecombe, *The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills.* Neuropsychology Review, 2003. **13**(4): p. 181-97.

71. Micceri, T., *The Unicorn, the Normal Curve, and Other Improbable Creatures.* Psychological Bulletin, 1989. **105**(1): p. 156-166.
72. Ruff, R.M., *A friendly critique of neuropsychology: facing the challenges of our future.* Archives of Clinical Neuropsychology, 2003. **18**(8): p. 847-64.
73. Nyberg, L. and J. Eriksson, *Working Memory: Maintenance, Updating, and the Realization of Intentions.* Cold Spring Harbor Perspectives in Biology, 2015. **8**(2): p. a021816.
74. Aben, B., S. Stapert, and A. Blokland, *About the Distinction between Working Memory and Short-Term Memory.* Front Psychol, 2012. **3**: p. 301.
75. Moser, M.B., D.C. Rowland, and E.I. Moser, *Place cells, grid cells, and memory.* Cold Spring Harbor Perspectives in Biology, 2015. **7**(2): p. a021808.
76. Cobos, B., K. Haskard-Zolnierek, and K. Howard, *White coat hypertension: improving the patient-health care practitioner relationship.* Psychol Res Behav Manag, 2015. **8**: p. 133-41.
77. Kazemi, R., et al., *Effects of Shift Work on Cognitive Performance, Sleep Quality, and Sleepiness among Petrochemical Control Room Operators.* J Circadian Rhythms, 2016. **14**: p. 1.
78. Kohs, S.C., *The Block-Design Tests.* Journal of Experimental Psychology, 1920. **5**(3): p. 357-376.
79. Dirks, J., *The Effect of a Commercial Game on Childrens Block Design Scores on the Wisc-R Iq Test.* Intelligence, 1982. **6**(2): p. 109-123.
80. Chelune, G.J. and K. Duff, *The Assessment of Change: Serial Assessments in Dementia Evaluations.* Handbook on the Neuropsychology of Aging and Dementia, 2013: p. 43-57.
81. Sternberg, R.J., *Encyclopedia of human intelligence*1994, New York; Toronto; New York: Macmillan ; Maxwell Macmillan Canada ; Maxwell Macmillan International.
82. Theisen, M.E., et al., *Effects of practice in repeated administrations of the Wechsler Memory Scale Revised in normal adults.* Assessment, 1998. **5**(1): p. 85-92.
83. Hassenstab, J., et al., *Absence of practice effects in preclinical Alzheimer's disease.* Neuropsychology, 2015. **29**(6): p. 940-8.
84. Rabbitt, P., et al., *Practice and drop-out effects during a 17-year longitudinal study of cognitive aging.* Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 2004. **59**(2): p. P84-97.

85. Yan, J.H. and M.B. Dick, *Practice effects on motor control in healthy seniors and patients with mild cognitive impairment and Alzheimer's disease.* Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition, 2006. **13**(3-4): p. 385-410.
86. Suchy, Y., M.L. Kraybill, and E. Franchow, *Practice effect and beyond: reaction to novelty as an independent predictor of cognitive decline among older adults.* Journal of the International Neuropsychological Society, 2011. **17**(1): p. 101-11.
87. Kruger, J. and D. Dunning, *Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments.* Journal of Personality and Social Psychology, 1999. **77**(6): p. 1121-34.
88. Schoo, L.A., et al., *Insight in cognition: self-awareness of performance across cognitive domains.* Appl Neuropsychol Adult, 2013. **20**(2): p. 95-102.
89. Geraci, L., et al., *The effect of cognitive testing and feedback on older adults' subjective age.* Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition, 2018. **25**(3): p. 333-350.
90. Eckerstrom, M., et al., *High Prevalence of Stress and Low Prevalence of Alzheimer Disease CSF Biomarkers in a Clinical Sample with Subjective Cognitive Impairment.* Dementia and Geriatric Cognitive Disorders, 2016. **42**(1-2): p. 93-105.
91. Fresson, M., B. Dardenne, and T. Meulemans, *Diagnosis threat and underperformance: The threat must be relevant and implicit.* Journal of Clinical and Experimental Neuropsychology, 2018. **40**(7): p. 682-697.
92. Fresson, M., et al., *[Formula: see text]The effect of stereotype threat on older people's clinical cognitive outcomes: investigating the moderating role of dementia worry.* Clin Neuropsychol, 2017. **31**(8): p. 1306-1328.
93. Wolfsgruber, S., et al., *Memory concerns, memory performance and risk of dementia in patients with mild cognitive impairment.* PLoS ONE, 2014. **9**(7): p. e100812.
94. Roberts, J.L., L. Clare, and R.T. Woods, *Subjective memory complaints and awareness of memory functioning in mild cognitive impairment: a systematic review.* Dementia and Geriatric Cognitive Disorders, 2009. **28**(2): p. 95-109.
95. Reisberg, B. and S. Gauthier, *Current evidence for subjective cognitive impairment (SCI) as the pre-mild cognitive impairment (MCI) stage of subsequently manifest Alzheimer's disease.* International Psychogeriatrics, 2008. **20**(1): p. 1-16.

96. Thames, A.D., et al., *Effects of stereotype threat, perceived discrimination, and examiner race on neuropsychological performance: simple as black and white?* Journal of the International Neuropsychological Society, 2013. **19**(5): p. 583-93.
97. Steele, C.M. and J. Aronson, *Stereotype threat and the intellectual test performance of African Americans.* Journal of Personality and Social Psychology, 1995. **69**(5): p. 797-811.
98. Wickes, T.A., *Examiner Influence in a Testing Situation.* Journal of Consulting Psychology, 1956. **20**(1): p. 23-26.
99. Hessen, E., *Rehearsal significantly improves immediate and delayed recall on the Rey Auditory Verbal Learning Test.* Applied Neuropsychology, 2011. **18**(4): p. 263-8.
100. Casas, R., et al., *Interpreter-mediated neuropsychological testing of monolingual Spanish speakers.* Clin Neuropsychol, 2012. **26**(1): p. 88-101.
101. Bialystok, E., F.I. Craik, and G. Luk, *Bilingualism: consequences for mind and brain.* Trends Cogn Sci, 2012. **16**(4): p. 240-50.
102. de Bruin, A., B. Treccani, and S. Della Sala, *Cognitive advantage in bilingualism: an example of publication bias?* Psychological Science, 2015. **26**(1): p. 99-107.
103. Stalhammar, J., A. Nordlund, and A. Wallin, *An example of exceptional practice effects in the verbal domain.* Neurocase, 2015. **21**(2): p. 162-8.
104. Stalhammar, J., et al., *Boston Naming Test automatic credits inflate scores of nonaphasic mild dementia patients.* Journal of Clinical and Experimental Neuropsychology, 2016. **38**(4): p. 381-92.
105. Nordlund, A., et al., *The Cognitive Assessment Battery (CAB): a rapid test of cognitive domains.* International Psychogeriatrics, 2011. **23**(7): p. 1144-51.
106. Kaplan, E., et al., *Boston Naming Test*2001: Pro-Ed, 8700 Shoal Creek Blvd, Austin, TX 78757-6897.
107. Lezak, P.N.P.N.M.D., et al., *Neuropsychological Assessment*. 4th ed2004, Oxford ; New York: Oxford University Press.
108. Yates, F.A., *The art of memory*1966, London,: Routledge & K. Paul.
109. Tallberg, I.M., *The Boston Naming Test in Swedish: normative data.* Brain and Language, 2005. **94**(1): p. 19-31.
110. Van der Elst, W., et al., *A large-scale cross-sectional and longitudinal study into the ecological validity of neuropsychological test measures in neurologically intact people.* Archives of Clinical Neuropsychology, 2008. **23**(7-8): p. 787-800.

111. Willers, I.F., M.L. Feldman, and R.F. Allegri, *Subclinical naming errors in mild cognitive impairment: A semantic deficit?* Dement Neuropsychol, 2008. **2**(3): p. 217-222.
112. Balthazar, M.L., F. Cendes, and B.P. Damasceno, *Semantic error patterns on the Boston Naming Test in normal aging, amnestic mild cognitive impairment, and mild Alzheimer's disease: is there semantic disruption?* Neuropsychology, 2008. **22**(6): p. 703-9.
113. Budd, M.A., *Boston Naming Test with Latencies (BNT-L)*, 2007, University of North Texas: Denton, Texas. UNT Digital Library.
114. Pedraza, O., et al., *Differential item functioning of the Boston Naming Test in cognitively normal African American and Caucasian older adults.* Journal of the International Neuropsychological Society, 2009. **15**(5): p. 758-68.
115. Sarilidou, T.P., *support mail, answer regarding BNT development*, J. Stålhammar, Editor 2019, Hogrefe Psykologiförlaget AB.
116. Lanehart, S.L., *The Oxford handbook of African American language*. Oxford handbooks. xxxvii, 903 pages.
117. Gustafsson, A.W.H., David, *Ord på prov. En studie av ordförståelse i högskoleprovet*. Lundastudier i nordisk språkvetenskap. Vol. Serie A 75. 2017.
118. Baddeley, A., *Working memory: theories, models, and controversies.* Annual Review of Psychology, 2012. **63**: p. 1-29.
119. Adesope OO, L.T., Thompson T, Ungerleider C, *A systematic review and meta-analysis of the cognitive correlates of bilingualism.* Review of Educational Research, 2010(80): p. 207–245.
120. Ljungberg, J.K., et al., *A longitudinal study of memory advantages in bilinguals.* PLoS ONE, 2013. **8**(9): p. e73029.
121. Dennett, D.C., *Consciousness explained / Daniel C. Dennett ; illustrated by Paul Weiner*1991, Boston: Little, Brown and Co.
122. Golestani, N., S. Rosen, and S.K. Scott, *Native-language benefit for understanding speech-in-noise: The contribution of semantics.* Biling (Camb Engl), 2009. **12**(3): p. 385-92.
123. Bosker, H.R. and E. Reinisch, *Foreign Languages Sound Fast: Evidence from Implicit Rate Normalization.* Front Psychol, 2017. **8**: p. 1063.
124. Gollan, T.H. and M. Goldrick, *Does bilingualism twist your tongue?* Cognition, 2012. **125**(3): p. 491-7.
125. Luk, G. and E. Bialystok, *Bilingualism is not a categorical variable: Interaction between language proficiency and usage.* J Cogn Psychol (Hove), 2013. **25**(5): p. 605-621.
126. Stern, Y., *Cognitive reserve in ageing and Alzheimer's disease.* Lancet Neurology, 2012. **11**(11): p. 1006-12.

127. Harada, C.N., M.C. Natelson Love, and K.L. Triebel, *Normal cognitive aging.* Clinics in Geriatric Medicine, 2013. **29**(4): p. 737-52.
128. Salthouse, T.A., *Implications of within-person variability in cognitive and neuropsychological functioning for the interpretation of change.* Neuropsychology, 2007. **21**(4): p. 401-11.
129. Almkvist, O., *Alzheimers's disease and related dementia disorders : neuropsychological identification, differentiation, and progression*, 1993, Diss. (sammanfattning) Stockholm : Karol. inst.: Stockholm.
130. Wu, Q., J.S. Chan, and J.H. Yan, *Mild cognitive impairment affects motor control and skill learning.* Reviews in the Neurosciences, 2016. **27**(2): p. 197-217.
131. Draaisma, D., *The Nostalgia factory : memory, time and ageing*2013, USA: Yale University Press. xiv, 158 pages.
132. Jansari, A. and A.J. Parkin, *Things that go bump in your life: explaining the reminiscence bump in autobiographical memory.* Psychology and Aging, 1996. **11**(1): p. 85-91.
133. Costa, P. and R. Kastenbaum, *Some aspects of memories and ambitions in centenarians.* Journal of Genetic Psychology, 1967. **110**(1st Half): p. 3-16.
134. Price, D.D., D.G. Finniss, and F. Benedetti, *A comprehensive review of the placebo effect: recent advances and current thought.* Annual Review of Psychology, 2008. **59**: p. 565-90.
135. Turi, Z., et al., *Evidence for Cognitive Placebo and Nocebo Effects in Healthy Individuals.* Sci Rep, 2018. **8**(1): p. 17443.
136. Hessen, E., et al., *Subjective Cognitive Impairment Is a Predominantly Benign Condition in Memory Clinic Patients Followed for 6 Years: The Gothenburg-Oslo MCI Study* . Dementia and Geriatric Cognitive Disorders Extra, 2017. **7**(1): p. 1-14.
137. Reisberg, B., et al., *Outcome over seven years of healthy adults with and without subjective cognitive impairment.* Alzheimers Dement, 2010. **6**(1): p. 11-24.
138. Miller, J.B. and W.B. Barr, *The Technology Crisis in Neuropsychology.* Archives of Clinical Neuropsychology, 2017. **32**(5): p. 541-554.
139. Peal, E. and W.E. Lambert, *The relation of bilingualism to intelligence*. Psychological monographs: general and applied,1962, Washington: American Psychological Association. 23 p.