Cognitive development and educational attainment across the life span

Rasmus Berggren

Karolinska Institutet

From Aging Research Center
Karolinska Institutet, Stockholm, Sweden

COGNITIVE DEVELOPMENT AND EDUCATIONAL ATTAINMENT ACROSS THE LIFE SPAN

Rasmus Berggren

Karolinska Institutet

Stockholm 2019

Cognitive development and educational attainment across the life span

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Rasmus Berggren

Principal Supervisor:
Prof. Martin Lövdén
Karolinska Institutet
Department of NVS
Aging Research Center

Co-supervisor(s):
Prof. Yvonne Brehmer
Tilburg University
Department of Developmental
Psychology
TS Social and Behavioral Sciences

Dr. Jonna Nilsson
Karolinska Institutet
Department of NVS
Aging Research Center

Opponent:
Dr. Stuart Ritchie
King's College London
Social, Genetic and Developmental
Psychiatry Centre

Examination Board:
Professor Carl-Johan Boraxbekk
Umeå University
Center of Demographic and
Aging Research

Professor Maria Larsson
Stockholm University
Department of Psychology

Dr. Valgeir Thorvaldsson
University of Gothenburg
Department of Psychology

ABSTRACT

In this thesis I explore two strands of research: in the first half I explore the effects of language learning on brain structure and cognitive function. In the second half I study to what extent educational attainment alter the rate of cognitive decline in old age. These different strands of research are united through the concept of brain plasticity, which is the brain's ability to change its structural configuration in response to new experiences.

In Study I we charted the neural underpinning of foreign language learning in a sample of younger adults. A total of 56 younger adults were randomized to either a 10-week beginner's course in Italian or a control condition. For those studying Italian we found that grey-matter change in the right hippocampus was associated with how much time they spent practicing, rather than with how good they became suggesting that effort, rather than achieved proficiency, is what drives neuroplastic change.

Study II presents the largest (to date) randomized trial looking into the causal effects of language learning as cognitive engagement in older age, specifically foreign language learning. 160 people between ages 65 and 75 were randomized to either an 11-week beginner's course in Italian, or an 11-week relaxation training course. While we predicted that language learning would improve cognitive function, specifically associative memory, we found no evidence to support this hypothesis.

Study III and IV address the question of whether or not educational attainment affects the rate of cognitive decline in old age. While it is clear that educational attainment is associated with level of cognitive function, we find no evidence that it alters the rate of decline in old age. Study III address this question using a novel statistical approach while Study IV presents a meta-analysis on the subject, arriving at similar conclusions.

LIST OF SCIENTIFIC PAPERS

I. Bellander, M., **Berggren, R.,** Mårtensson, J., Brehmer, Y., Wenger, E., Li, TQ., Bodammer, NC., Shing, YL., Werkle-Bergner, M., Lövdén, M. (2016). Behavioral correlates of changes in hippocampal gray matter structure during acquisition of foreign vocabulary. *NeuroImage*, 131(205-213)

II. **Berggren, R.,** Nilsson, J., Brehmer, Y., Schmiedek, F., Lövdén, M. (Under review). Foreign language learning in older age does not improve memory or intelligence: Evidence from a randomized controlled study.

III. **Berggren, R.,** Nilsson, J., Lövdén, M. (2018). Education does not affect cognitive decline in aging: A Bayesian assessment of the association between education and change in cognitive performance. *Frontiers in Psychology*, 9:1138.

IV. Seblova, D., **Berggren, R.,** Lövdén, M. (Manuscript). Education and age-related decline in episodic memory performance: systematic review and meta-analysis of longitudinal studies.

CONTENTS

1   INTRODUCTION

Decline of cognitive functioning is a hallmark symptom of aging, and while most experience some alteration in old age not everyone is affected equally. Pathological manifestations of cognitive impairment, such as Alzheimer's disease, severely limit the independence and capabilities of the afflicted patients, but age-related cognitive decline is also a natural consequence of healthy aging. As this process may start long before any manifest symptoms occur it is important to understand what factors predict age-related cognitive decline in order to identify those at risk. Another question is if, or rather what we can do about this process once it has started, i.e. if the process is at all reversible, and if so, how?

1.1   Cognitive functioning across the life span

The umbrella term 'cognitive functions' refers to brain-based behavioral abilities we use to carry out everyday tasks. Of focal interest for the present thesis are fluid intelligence, crystallized intelligence, episodic memory, associative memory and working memory. Fluid intelligence refers to reasoning, abstract thinking, and pattern recognition (e.g. the ability to see similarities between distinct phenomena), and is involved in solving problems with which we have little or no prior experience. Fluid intelligence is also stable predictor of educational success (Rohde & Thompson, 2007), mortality risk (Aichele, Rabbitt, & Ghisletta, 2015) and everyday functioning in old age (Tucker-Drob, 2011) and exhibits strong age-related decline beginning in early adulthood (Rönnlund, Nyberg, Bäckman, & Nilsson, 2005; Salthouse, 2009). Crystallized intelligence refers to the ability to apply already acquired knowledge in familiar situations, such as factual knowledge and reading comprehension (Cattell, 1963). Crystallized intelligence is strongly associated with educational attainment (Gerstorf, Ram, Hoppmann, Willis, & Schaie, 2011) and, in contrast to fluid intelligence, remain relatively preserved even into old age (Rönnlund, Nyberg, Bäckman, & Nilsson, 2005). According to Cattell's investment theory of intelligence (Cattell, 1963) fluid intelligence is invested early on in order to acquire skills and abilities, which subsequently become crystallized. One example of such an ability is reading comprehension. Working memory refers to the limited capacity to maintain, manipulate and update information in short-term memory (Baddeley, 1992), such as remembering a telephone number. Like fluid intelligence, working memory begin to decline in early adulthood. Episodic memory refers to the encoding and retrieval of declarative memories with spatio-temporal characteristics, such as autobiographical memories (Tulving, 1972). Episodic memory is a function of two distinct memory systems: item memory, involving the recollection of single items (e.g. a face or a name), and associative memory, involved in the binding of two distinct pieces of information (e.g. the pairing of a name with a face; Davachi, 2006). Episodic memory also exhibits marked age-related decline (Rönnlund et al., 2005). Associative memory also

exhibits marked age-related decline, whereas item memory is relatively preserved even in older age. Thus, age-related decline in episodic memory may be due to a selective deficiency in associative memory (Naveh-Benjamin, 2000).

Cognitive aging has been studied using both cross-sectional designs, in which participants of different ages are measured at one point in time, and longitudinal study designs, in which participants are measured repeatedly across multiple time points (Baltes, 1968; Schaie, 1965). Results from cross-sectional and longitudinal studies on cognitive aging suggest somewhat different patterns of decline (Salthouse, 2009; Schaie, 1994, 2005). Generally, cross-sectional studies suggest earlier onset and more rapid decline than longitudinal studies (Schaie, 1994). This can be partly explained by differences in design: while cross-sectional designs measure "change" as between-subject (interindividual) differences, longitudinal designs measure "change" as within-subject (intraindividual) differences, i.e. change over time. For example, observed age-related differences in cross-sectional research may partly reflect true age-related changes (i.e. changes that reflect the biological process of aging), but may also be confounded by cohort effects (e.g. effects applying to everyone at the same time) or period effects (i.e. effects applying to everyone present at a point in time, such as wars, famine or changes in policy). Another source of variability is the so-called Flynn effect (Flynn, 1984), the fact that later generations tend to score higher than earlier generations on cognitive tests (Rönnlund & Nilsson, 2009). Proposed explanations for this include reduced family-size (Sundet, Eriksen, & Tambs, 2008), increased nutritional quality (Colom, Lluis-font, & Andre, 2005; Lynn, 1990), better access to and higher quality of schooling (Ceci, 1991; Lager, Seblova, & Falkstedt, 2016; Ritchie & Tucker-drob, 2018), and an increased familiarity with cognitive testing (Neisser, 1997). This might reflect societal changes and improved living conditions.

Teasing apart the different influences of age, period and cohort is a fundamental problem in longitudinal aging research (Schaie, 2007). They cannot all be estimated simultaneously due to multicollinearity, i.e. one factor can be completely determined by the two others (for example, age = cohort + period). On the other hand, longitudinal designs might underestimate the true age-related change, e.g. due to learning effects of repeated testing (re-test effects; Salthouse, 2009), or suffer from selective attrition (e.g. that a continued participation in the study depends on the outcome of interest, such as health status or cognitive performance) which might produce biased estimates. The two designs can be fruitfully combined, e.g. by a cohort-sequential longitudinal designs in which multiple cohorts are followed over time while new cohorts are introduced at each new time point (Nilsson et al., 1997; Schaie, 1965). These combine the relative advantages while offsetting the relative weaknesses of cross-sectional and longitudinal designs.

A person's cognitive performance can be described as a function of initial performance at some time point (level) and change thereafter (slope). Consequently, it is

important to consider both factors associated with level and factors associated with slope (Ritchie et al., 2016) as both may contribute to the development of cognitive impairment: those with higher initial cognitive functioning will take longer to reach an impairment threshold, as will those with a slower rate of decline. For example, educational attainment has an established association with level of cognitive functioning but not with rate of decline (Lenehan, Summers, Saunders, Summers, & Vickers, 2014; Zahodne et al., 2011) while both engagement in leisure activities and occupational complexity are associated both with level of cognitive functioning and rate of decline (Finkel, Andel, Gatz, & Pedersen, 2009; Köhncke et al., 2016).

1.2   Educational attainment and cognitive functioning

Formal schooling begins at an early age and extends through late adolescence, sometimes into early adulthood. For most of us school is, among other things, a place to learn new things. Educational attainment is also a stable predictor of a wide range of life outcomes. Of particular importance for the present thesis is the association between educational attainment and cognitive functioning. Educational attainment has a well-established association with level of cognitive functioning (e.g. Opdebeeck, Martyr, & Clare, 2015; Strenze, 2007) but the nature of this relationship is unclear. One possibility is that subjects with higher innate cognitive ability go on to pursue higher education (Kremen et al., 2019). Another possibility is that education has a causal influence on cognitive function, such that staying in school for longer increase cognitive function (Ceci, 1991; Ritchie & Tucker-drob, 2018). A third option is that the associations are explained by other variables associated with both intelligence and educational attainment, e.g. parental socioeconomic status. One may also speculate that the relationship between education and cognition changes over time, e.g. in response to societal changes or policy implementation. For example, the introduction or increase in mandatory schooling may truncate the distribution on the lower end of the spectrum, e.g. people who would have otherwise dropped out of school will stay in school up to the minimum required duration, even if more education does not bring additional cognitive benefits, thus obscuring the underlying relationship. It may also be that wider access to education triggers reactive changes, such as altered requirements and altered educational content. This creates complex interactions between education, cohort effects and period effects, and which one of these scenarios – if any – is correct may have implications for shaping policy; if education has no causal effect on cognitive function then increasing mandatory schooling will not result in higher intelligence.

Discerning the nature and direction of causal relationships is therefore of outmost importance. While conducting randomized controlled trials (RCTs) are commonly viewed as the golden standard for causal inference, certain areas – such as educational policy – do not easily lend themselves to tightly controlled experiments. To circumvent this researcher may exploit natural variation in exposure, such as successive implementation of policy changes in different areas, to arrive at causal

conclusions. Using this design, Lager et al. (2016) showed that increasing mandatory schooling by one year increased full-scale IQ by 0.75 points. Similarly, a meta-analysis by Ritchie & Tucker-Drob (2018) found that a one-year increase in schooling increased IQ by between 1 and 5 points.

Educational attainment is also a central component of the concept of cognitive reserve (Stern, 2002, 2010). Cognitive reserve has been introduced to account for often observed discrepancy between predicted cognitive function, as expected from neuroanatomical and – histological samples, and the manifest cognitive function as assessed by psychological testing. Cognitive reserve comprises at least three distinct theoretical models. A passive reserve model suggest that different brains have different brain reserve capacity, manifested as brain volume or synapse count, and that brains with more brain reserve capacity are simply able to withstand more damage before exhibiting cognitive impairment (Stern, 2002). According to an active reserve model, the brain copes with brain injury through more efficient use of neural resources (cognitive reserve) or by differential recruitment of brain networks (compensation).

These three models make different predictions regarding the role of educational attainment and cognitive decline. The active reserve model imply that higher educational attainment is associated with slower decline; the compensatory model suggest that higher educational attainment is associated with faster decline (but perhaps later onset); and the passive model suggest that education only has an effect on level of cognitive function but is unrelated to cognitive decline (Lenehan et al., 2014). A narrative review by Lenehan et al. (2014) found somewhat conflicting evidence but generally favored the notion of passive cognitive reserve. Studies reporting either positive of negative effects of education on rate of decline suffer from methodological drawbacks in that they tend to use less sophisticated statistical methods, have shorter follow-up periods or comprise samples of relatively low average educational attainment. Study III and Study IV in this thesis further investigate the role of educational attainment on cognitive decline.

1.3   Brain plasticity

Brain plasticity refers to "the ability of the brain to change its structure in response to a mismatch between its capacity and the environmental demands" (Lövdén, Bäckman, Lindenberger, Schaefer, & Schmiedek, 2010). Within this framework plasticity denotes an increase in capacity, i.e. an increase in the range of functional supply. This can be distinguished from flexibility, which is performance variability within the limits of the functional range. A requirement for plastic changes is that they are reactive changes, in response to altered demands, and that they are associated with structural changes in the brain leading to increases in capacity. Flexible changes, on the other hand, do not require structural change but may be associated with mere functional changes, e.g. changes in neural activity.

A paradigmatic case of plastic change can be the brain's response following brain injury. A stroke may disrupt neuronal tissue, which is followed by an immediate reduction the functional behavioral range (the capacity). The result is fewer available resources to meet demands which in turn puts a bigger stress on the system to carry out the same task. A sustained discrepancy between supply and demand may induce plastic changes, which over time may lead to structural changes, e.g. rewiring of white-matter tracts and volumetric changes in grey matter. Another example of brain plasticity is structural rewiring such as long-term potentiation (LTP) whereby synaptic connections between neuronal clusters are strengthened through repeated firing in response to a stimulus. Plastic changes is not expected to occur if the environmental demands lie within the current functional supply (i.e. if the organism can respond with existing flexibility) nor are they expected to occur if the environmental demands lie too far outside the current functional supply. In other words, the environmental demands cannot be too high nor too low, but must lie on the border of the current functional capacity in order to induce the necessary mismatch between capacity and demand to trigger plastic changes. This serves as a theoretical motivation behind adaptive cognitive training regimes, where task difficulty is altered depending on performance in order for the task to be suitably challenging.

Thus, plastic changes may occur during skill acquisition and learning. But if learning leads to macroscopic structural changes in the brain, e.g. increases in grey-matter volume, how can the confines of a severely limited space, such as the human skull, allow for virtually infinite learning? The renormalization-expansion model of brain plasticity (Wenger, Brozzoli, Lindenberger, & Lövdén, 2017) accounts for this conundrum by applying Darwinian principles of evolution and selection to the developing brain. According to this model, plastic changes happen during different phases of expansion, selection and renormalization. When acquiring a new skill, a wide range of candidates (e.g. neurons, glial cells, synapses) are recruited and tested for their functional suitability in carrying out the proposed task. This over-recruitment leads to expansion of neural tissue, such as increases in cortical thickness or grey-matter volume. Next, there is a selection among all these possible candidates, where the most suitable candidates are developed further (e.g. axonal growth, dendritic branching) and maintained (e.g. through neurotrophic factors). Finally, unsuitable candidates wither and atrophy, leading to subsequent renormalization of neural tissue.

A common method to study the human brain is through magnetic resonance imaging (MRI). MRI is a non-invasive method used to obtain spatiotemporal (3D or 4D) images of human brain tissue, such as grey- or white matter, and has been used to study experience-dependent changes in the human brain. In a seminal study, Draganski et al. (2004) demonstrated increases in grey-matter volume in the medial temporal lobe following three months of juggling training. In a group of 14 military interpreters, (Mårtensson et al. (2012) observed grey-matter increases

in the right hippocampus following three months of intense language studies. They also showed that structural growth in the middle frontal gyrus correlated with the amount of struggle experienced during language acquisition, suggesting plastic changes induced by prolonged mismatch between functional supply and environmental demands.

1.4   Language learning and cognitive training

Longitudinal studies have established an association between cognitive functioning and engagement in cognitively stimulating activities in old age (Valenzuela & Sachdev, 2009; Köhncke et al., 2016). This raises prospect of improving cognitive function by engaging in cognitively stimulating activities, such as cognitive training. Adaptive cognitive training paradigms involve effortful practice on a task specifically taxing one cognitive ability, such as a numerical updating task taxing working memory capacity, in which task difficulty is adjusted to match the participants' performance, so as to always lie at the upper limit of the participants' ability. Of particular interest is if and how improvements on the trained task (e.g. numerical updating) generalize to non-trained tasks (e.g. N-back) involving the same cognitive ability ("near-transfer") and how they generalize to non-trained tasks (e.g. matrix reasoning) involving a different cognitive ability ("far-transfer"). Strategy-based training, which involves developing successful strategies to solving the particular task at hand, are not expected to exhibit far transfer to other domains while process-based training, which aims to improve the cognitive ability per se (e.g. improving working memory capacity), are. While the prospect of improving broad cognitive functions through deliberate practice might be intuitive, experimental evidence do not unanimously support this hypothesis. There is an on-going debate regarding the efficacy of cognitive training in facilitating far-transfer. So far, the evidence for far-transfer in older adults has been limited (Karbach & Verhaeghen, 2014; Melby-Lervåg & Hulme, 2013).

Like cognitive training, foreign language learning is a cognitively challenging activity that has been proposed as another way of ameliorating cognitive decline in old age (Antoniou, Gunasekera, & Wong, 2013). Learning a foreign language involves engaging a wide range of cognitive abilities that are shown to be negatively affected by aging (Antoniou et al., 2013; Raz et al., 2005): working memory is needed to keep and manipulate semantic information in short term memory, e.g. in applying new grammatical rules and forming sentences; inductive reasoning is involved in extending previously learnt rules to new cases; task switching in switching between the native tongue and the foreign language; and associative memory is involved when acquiring a foreign vocabulary. The last component of associative memory is what we utilize in Study I and II, where we model the acquisition of a foreign vocabulary as associative memory training, in which a familiar word in the speaker's native tongue is repeatedly paired with a novel word in the target language (Breitenstein et al., 2005).

Another line of argument comes from the observational evidence bilinguals tend to out-perform monolinguals on executive functioning measures, such as inhibition and task switching (Bialystok, Craik, & Luk, 2012). The proposed explanation for better executive functioning among bilinguals is that bilinguals have a higher cognitive load than monolinguals because they need to suppress one language in favor of the other (inhibition) and be able to quickly switch between the two languages (task switching). This added cognitive load would accumulate over long periods of time, resulting in better cognitive functioning in old age – essentially exerting a natural instantiation of low-intensity cognitive training. However, the theory of bilingual advantage has been criticized lately (Paap, Johnson, & Sawi, 2015) arguing that observed benefits for bilinguals are a result of suboptimal statistical analyses or failure to control for important confounds. Further, it is also not obvious how any potential advantages from bilingualism would apply to foreign language learning in older age, unless such proficiency in the foreign language is acquired that it becomes a serious contender and to compete with the native language, in order to induce the executive conflict. With this in mind, we are unsure of the validity of the literature on bilingual advantage, and question the relevance of it to foreign language learning in old age.

Instead, we approach foreign language learning from a cognitive training perspective, specifically involving associative memory training. In contrast to item memory, which is relatively preserved in old age, associative memory shows marked age-related cognitive decline beginning around age 60 (Rönnlund et al., 2005). The hippocampus, located in the medial temporal lobe, is a neural structure crucial for the successful encoding and retrieval of episodic memories (Burgess, Maguire, & Keefe, 2002; Penfield & Milner, 1958), and in particular in the binding of associative memories, such as during foreign language acquisition (Breitenstein et al., 2005). Hippocampus also exhibits substantial deterioration with increasing age (Raz et al., 2005) and is a potential candidate where neurogenesis might occur (Kempermann, Song, & Gage, 2015). The exact relationship between associative memory and hippocampal grey-matter volume in older age is unclear but some findings suggest that episodic memory functioning has an age gradient, in that older people rely more on frontal regions for memory encoding and retrieval than younger people, who rely more on hippocampal regions (Becker et al., 2015; Becker, Kalpouzos, & Salami, 2017). Within the framework of neuroplasticity suggested by Lövdén et al. (2010) one could speculate that foreign vocabulary acquisition through associative memory training would induce the required mismatch between functional supply and environmental demands for neuroplastic changes in the hippocampus to occur. This is the theoretical basis for the language learning interventions in Study I and II.

2 AIMS

The present thesis contains two strands of research. The first half, studies I and II, are experimental studies investigating the effects of second-language learning in younger and older adults. The second part, studies III and IV, study cognitive decline across the life span and the moderating effects of educational attainment on the rate of decline. The concept of brain plasticity provides the theoretical framework against which this research can be understood.

3 SUMMARY OF STUDIES

3.1 Study I – Behavioral correlates of changes in hippocampal gray matter structure during acquisition of foreign vocabulary

Introduction. Human brain structure can change in response to altered environmental demands. Earlier research has demonstrated growth in grey matter volume following motor skill coordination training such as juggling (Draganski et al., 2004) and growth in hippocampal grey matter volume following three months of intense language learning among military interpreters (Mårtensson et al., 2012) but it is unclear whether volumetric increases reflect achieved proficiency or duration of engagement. Acquisition of a foreign vocabulary can be modeled on an associative process whereby a novel word in the foreign language is repeatedly paired with a word in the mother tongue. As associative memory crucially involves the hippocampus we predict that changes in hippocampal grey-matter volume predict foreign language acquisition.

Sample. 80 younger adults between ages 18-30 were recruited through ads in a local newspaper. 56 participants were randomized to a 10-week language learning course and 26 participants were randomized to an active control condition (watching Italian movies with Swedish subtitles). Participants were required to have no or minimal knowledge of any Roman languages. A total of 33 participants (61%) completed the language course and 23 participants (89%) completed the control condition.

Materials and methods. Participants in the language course met in a group setting for 2.5 hours, once per week for 10 weeks. During classes participants focused on verbal communication exercises. Between classes, participants were instructed to practice vocabulary learning using an iOS app designed for this study consisting of words from the course textbook. Participants in the control condition met once per week for 10 weeks and watched Italian movies with Swedish subtitles. Participants were tested using a computerized cognitive battery involving an associative memory task and a pattern separation task (delayed match-to-sample; DMS). At post-test, participants in the language learning condition did a vocabulary test of 100 words randomly selected from the app corpus. We also extracted the time spent on vocabulary acquisition from the app. All participants underwent structural and functional MRI before and after the intervention.

Statistical analysis. Behavioral data was analyzed using a 2 (group) x 2 (time) mixed ANOVA to detect changes in DMS or associative memory performance. Grey-matter volume in the language learning group was extracted using a paired t-test,

which was then entered in an exploratory path model to determine the relationship between grey-matter volume, baseline characteristics, and training-related gains.

Results. A whole-brain analysis revealed training-related grey-matter increases in the right hippocampus and the left occipital lobe for the language group. A group x time analysis suggested that this was specific to the language group. Average grey-matter volumetric increase was extracted from the hippocampal region-of-interest and entered into an exploratory path model, predicted by time spent training and acquired vocabulary. A first model suggested that time spent training was associated with vocabulary acquisition and hippocampal grey-matter change, but that acquired vocabulary was not associated with hippocampal grey-matter change (Fig 1). In a second model, we included associative memory and DMS baseline performance as predictors of time spent training, acquired vocabulary and grey-matter change. In this model, DMS was associated with associative memory performance, grey-matter change and time spent training.
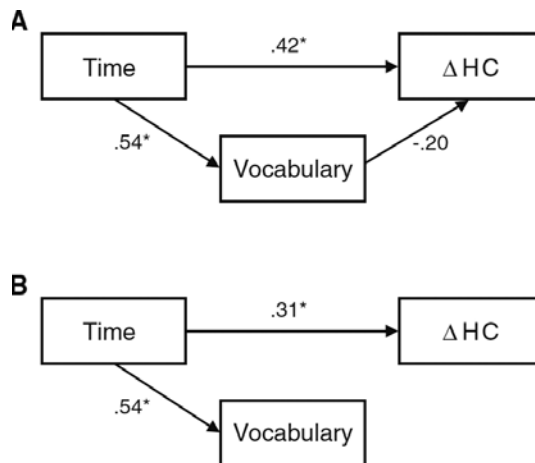
Figur 1. The basic path model of the relationship between time spent studying, vocabulary test score, and GM change in the right HC, (A) full model, and (B) pruned model.

Discussion. These findings align with previous findings on language acquisition in younger adults (e.g. Mårtensson et al., 2012), albeit in a less select sample. Time spent training exhibits a stronger association with hippocampal grey-matter increase than do vocabulary acquisition, suggesting that volumetric increases reflect effort rather than acquired skill. However, controlling for baseline cognitive functioning shows that only pattern separation ability uniquely predicts grey-matter change.

3.2 Study II – Foreign language learning in older age does not improve memory or intelligence: Evidence from a randomized controlled study

Introduction. Engagement in cognitively challenging activities has shown to be associated with better cognitive functioning in old age but experimental evidence on this association is lacking (Simons et al., 2016). Brain regions engaged in foreign language acquisition has shown substantial overlap with brain regions exhibiting age-related decline (Raz et al., 2005). Thus, foreign language learning has been proposed as a method to combat cognitive decline in old age.

Sample. 169 participants between ages 65–75 were recruited through ads in a local newspaper. A total of 160 participants completed the intervention. Participants were randomized into either an experimental group (N = 90) or a control group (N = 70) stratified on sex, age and baseline associative memory performance.

Materials and methods. Participants in the language learning condition met twice per week for 5 hours each week, for eleven weeks (totaling of 55 hours). During classes participants engaged in verbal communication exercises while following a textbook at a pace of about one chapter (2-4 pages) per week. Between classes participants practiced vocabulary acquisition. Participants in the relaxation training condition met once per week for 1 hour, for a total of 11 hours. During classes they focused on breathing and relaxation exercises.

Before and after the intervention participants did a cognitive testing battery with 3 tests of verbal intelligence (analogies, syllogisms and verbal inference), 2 tests of spatial intelligence (Raven's matrices and WASI matrices), 2 tests of working memory (numerical updating and N-back), 3 tests of associative memory (word-word, face-name and picture-picture) and 3 tests of item memory (word, faces, pictures). The tests were identical at pre- and post-test. At post-test, the language learning group was also given a vocabulary test of 110 words, randomly selected from the weekly word lists, to assess their acquired vocabulary.

A subsample also underwent structural and functional MRI (not reported here). Statistical analysis. The hypothesis was that language learning group would show larger pre-post gains than the relaxation group. We tested this hypothesis using both latent variable modeling and Bayesian model selection. Latent factors were formed of each test assumed to measure the same construct. Measurement invariance was tested by imposing increasingly strict constraints on factor loadings (weak MI), item intercepts (strong MI) and residual variances (strict MI). We estimated a latent change score model for both treatment groups simultaneously. The hypothesis of differential change was tested by comparing a model where the average change was freely estimated in both groups with a model where average

change was constrained to be equal in both groups. We also performed a Bayesian linear mixed regression on each observed measure, using Time (Pre vs Post) and Group (Language vs relaxation) and their interaction as fixed effects and including a random intercept. Bayesian hypothesis testing was done on the Time x Group interaction term, using informed half-normal priors.

Results. The vocabulary test showed that participants in the language group scored on average 57 words correct suggesting that they did acquire a basic vocabulary during the course. All latent factors passed the test of strict measurement invariance both in terms of absolute and relative fit. While both groups showed pre-post improvements, constraining the average change to equality across groups did result in significantly worse fit for any of the cognitive domains. This shows no greater improvement for the language learning group, and suggest that potential observed gains were due to re-test effects (Fig 2). This finding was further corroborated by Bayesian hypothesis testing weakly favoring the null hypothesis for all but one of the manifest variables.
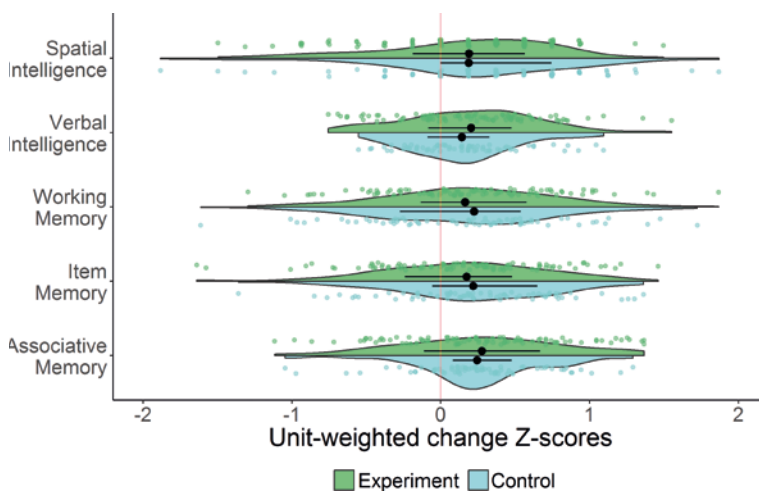
Figur 2. *Distribution of unit-weighted change (posttest – pretest) scores. Green dots denote individual change scores of participants in the language group and blue dots denote change scores for the participants in the relaxation group. Black dots and line segments denote the median and 1st and 3rd quartiles.*

Discussion. Participants in the language learning group did obtain a basic knowledge in the foreign language. We did not find evidence for the hypothesis that taking a beginner's course in a foreign language in older age improves verbal intelligence, spatial intelligence, working memory, associative memory or item memory. All latent factors exhibited acceptable psychometric properties and our sample size was adequate to detect relatively small differences. While the duration of the

course was more intense than a typical beginner's course and included additional vocabulary training, it is certainly possible that extended duration or more intense practice is necessary in order for language learning to translate into more general cognitive benefits. This study is also limited in that we only assess vocabulary and no other aspects (e.g. speech comprehension, speech production, grammar) of foreign language acquisition. It may also be that language learning is beneficial for certain target groups, such as those with low education or little cognitive stimulation. Overall, these results align with findings from similar studies and meta-analyses of the literature suggesting fairly limited general benefits of cognitive training.

3.3 Study III – Education does not affect cognitive decline in aging: a Bayesian assessment of the association between education and change in cognitive performance

Introduction. Cognitive performance decline in aging. Longitudinal studies suggest that decline in fluid abilities begin in early adulthood and episodic memory decline in late middle life while crystallized abilities remain relatively preserved even into old age (Rönnlund et al., 2005; Schaie, 1994). Educational attainment has a positive association with level of cognitive functioning but the association between educational attainment and rate of cognitive decline remains contested. Previous studies on the association between educational attainment and rate of decline report mixed findings (Lenehan et al., 2014). One methodological obstacle is the application of frequentist methods to assess situations where the null hypothesis might be plausible, as in the present case. To address this, we employed Bayesian hypothesis testing using informed priors from previous literature. We also explored to what extent the effect of education differs across cohorts.

Sample. This study uses data from the longitudinal cohort-sequential Betula study (Nilsson, 1997). In this study we sampled 1707 participants (54% female) aged 35-80 years with an average educational attainment of 10.2 years. Participants were tested on visuospatial ability, semantic knowledge and episodic memory at each measurement occasion.

Statistical analysis. We specified a Bayesian linear mixed model with sex, education and cohort as between-subject variables and age as a within-subject variable. We also included random intercepts and random linear slopes. Bayesian hypothesis testing requires an explicit specification of the alternative hypothesis. To this end we obtained estimates from two previous target articles the similar cognitive domains, similar age-range and using similar standardization, which facilitate the formulation of informed priors. The prior for the focal parameter, the age × education interaction term, is specified as normally distributed around 0 with a variance equal to 1, 2 or 4 times the observed estimates from the two target articles.

Bayesian hypothesis testing is done using the Savage-Dickey method. Inference regarding the cohort × education interaction is done using maximum a posteriori estimates and 95% highest density intervals.

Results. Results from the linear mixed model show that higher educational attainment is associated with greater performance on visuospatial ability, semantic knowledge and episodic memory but that educational attainment do not alter the rate of decline in either of these abilities, as indicated by parallel slopes (Fig 3). Contrary to expectation, our results suggest that the effect of education on crystallized abilities might decrease for later cohorts.

Discussion. We replicate the association between educational attainment and level of cognitive functioning. This association was stronger for semantic knowledge
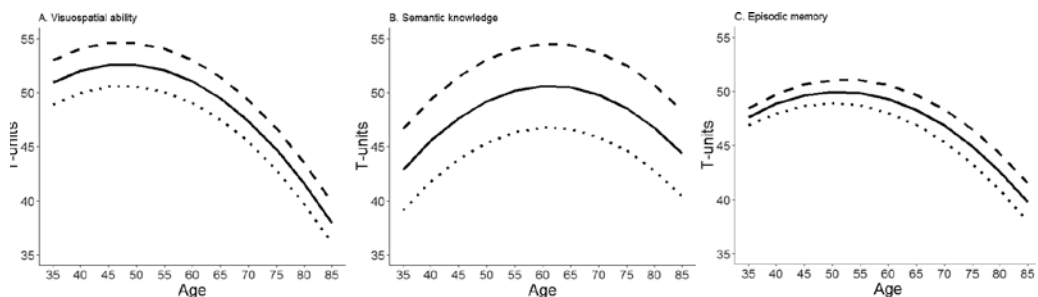
Figure 3. *Model-implied growth trajectories for (A) visuospatial ability, (B) semantic knowledge, and (C) episodic memory. Trajectories apply to men born in 1935 with 3 years more than average education (dashed line), average education (solid line), and 3 years less than average education (dotted line).*

than for visuospatial ability or episodic memory. Hypothesis testing on the association between educational attainment and rate of decline yields Bayes factors close to 1 for visuospatial ability and semantic knowledge, suggesting the data is unable to discriminate between the null and the alternative hypothesis. In retrospect, this is unsurprising given the precision of the original articles, suggesting that our findings should not alter the conclusions from previous articles, namely that education had a negligible effect on rate of decline. For episodic memory we find slight evidence in favor of the null hypothesis, with more ambitious alternative hypotheses yielding stronger evidence for the null hypothesis of no effect.

3.4 Study IV – Education and age-related decline in episodic memory performance: systematic review and meta-analysis of longitudinal studies

Introduction. Educational attainment has an established relationship with level of cognitive functioning but studies on the association between educational attainment and rate of cognitive decline has reported mixed evidence. In this study we perform a meta-analysis of 15 articles reporting on the association between educational attainment and rate of episodic memory decline.

Methods. We searched databases on articles reporting on the association between educational attainment and rate of episodic memory decline. From relevant articles we transformed each reported estimate to the association between 1 additional year of education and episodic memory change (in SDs) per decade, rescaled by baseline SD. We performed an inverse-variance weighted random-effects meta-analysis. In a subsequent meta-regression we also investigated the moderating effects of age, educational attainment or follow-up period. We performed a sensitivity analysis by removing outliers with large effect sizes and/or standard errors. Heterogeneity was assessed using Cochran's Q and I^2.

Results. The systematic search identified 15 articles reporting a total of 35 estimates on the association between educational attainment and rate of episodic memory decline. The total sample size was N = 92 930. The meta-analytic estimate was small and non-significant (β = 0.0021, p = .58), thus not supporting the notion that education affects rate of episodic memory decline. Average age, educational attainment or follow-up period did not change this conclusion, neither did removal of outliers. However, there was considerable heterogeneity across studies.
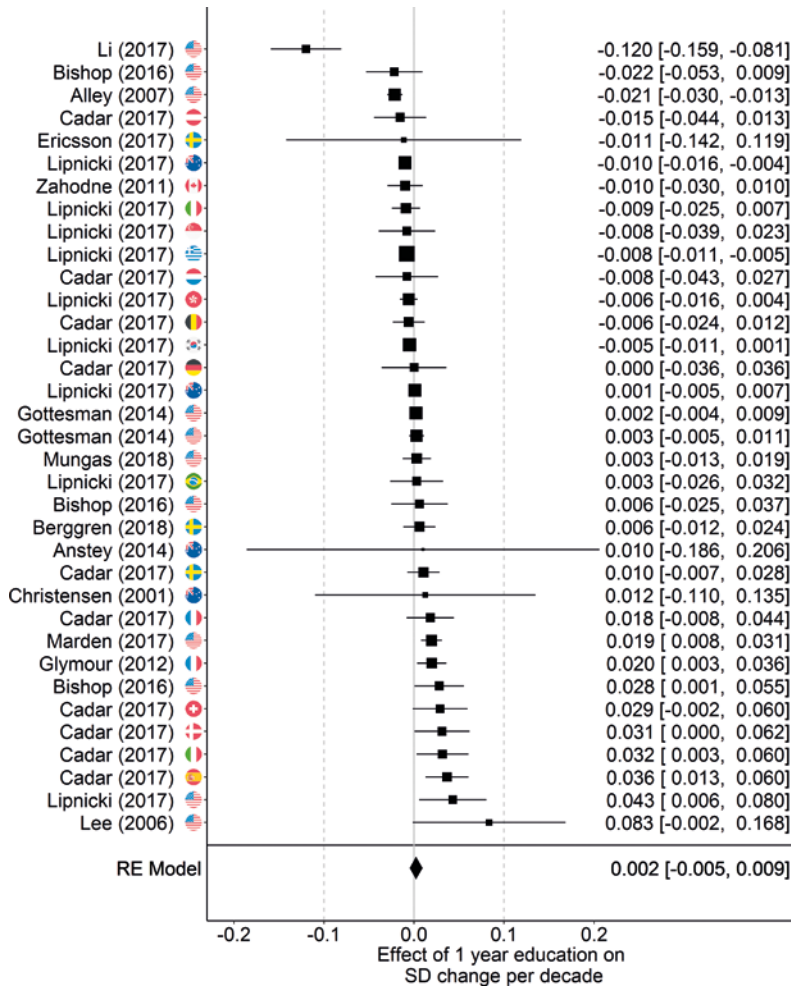
Figure 4. *Forest plot of episodic memory meta-analysis, point estimates and 95% confidence intervals. Size of points correspond to weight in meta-analysis.*

Discussion. This quantitative meta-analysis converges with other review articles (e.g. Lenehan et al., 2014) suggesting no or negligible effect of educational attainment on rate of episodic memory decline in old age. This result seems consistent over different tiers of educational attainment, across different ages and does not depend on length of follow-up period. While education has a substantial association with level of cognitive function, e.g. episodic memory, its effect on rate of decline is negligible, meaning we should perhaps look for other causes of individual differences in rate of cognitive decline.

4 DISCUSSION

4.1 Summary of results

The first half of this thesis presents findings from two experimental studies: one study on the neural correlates of language acquisition in younger adults (Study I) and one study on the cognitive benefits of foreign language learning in older adults (Study II). The second half of the thesis contains findings on the association between educational attainment and rate of decline: one original longitudinal study (Study III) and one meta-analysis (Study IV). Study I found that grey-matter volume increases in the right hippocampus was greater for younger adults participating in an eight-week language learning course compared to those partaking in an active control condition. Further, exploratory path analyses suggested that participant pattern separation ability at baseline significantly predicted grey-matter volume increase, controlling for time spent training on a vocabulary acquisition task. Study II found that eleven weeks of similar language training, with an emphasis on vocabulary acquisition, in older adults did not improve neither associative memory, item memory, working memory, spatial intelligence nor verbal intelligence compared to an active control group. Study III found that educational attainment did not alter the linear rate of cognitive decline in either episodic memory, visuospatial reasoning or semantic knowledge. In a similar vein, Study IV found that the meta-analytic estimate of educational attainment on rate of episodic memory decline is negligible.

4.2 Brain plasticity

In Study I we investigated the neural correlates associated with foreign language learning and found that grey-matter increase in the right hippocampus correlated with time investment, but not with acquired vocabulary. Our interpretation is that this grey-matter change might reflect plastic changes, but understanding the temporal dynamics of plastic changes at various stages is of course crucial. Wenger et al. (2016) has provided some experimental evidence for the temporal dynamics of plastic changes following motor skill learning. Their findings suggest that plastic changes occur relatively quickly and are fairly transient, occurring over the course of a few weeks, after which renormalization begins. While we are unable to address the temporal dynamics of plastic changes using a simple pre-post design, we can speculate that renormalization had been on-going for some time at the time of the second measurement occasion. It should also be noted that the findings of Wenger et al. (2016) apply to motor skill learning, and that the same pattern might not apply to higher cognitive processes, such as language learning.

More generally, it may be that plastic changes are most prominent during initial phases of skill acquisition, which is arguably when the most rapid learning occurs. During later stages it might be more a matter of honing skills, which might involve

functional changes rather than macroscopic structural changes, and that once the learning curve "levels off" renormalization begins. Furthermore, for complex tasks involving multiple cognitive functions, such as language learning, plastic changes may occur in different brain regions at different times.

4.3   Future prospects for cognitive training

In Study II we performed a relatively well-powered randomized controlled trial investigating the cognitive benefits of foreign language learning in old age. As we modelled foreign language acquisition as associative memory training, the most plausible benefit would be improved associative memory performance among foreign language learners. However, our findings do not support the hypothesis that participating in a moderately intense foreign language learning course improves neither memory nor intelligence, at least not in the short-term.

Proponents of language learning as cognitive engagement may point out that the observed lack of cognitive benefit was a result of insufficient intensity or duration, e.g. counterfactual claims of the sort "Had the course been longer (or more intense) then we might have seen improvements". This is, in a sense, necessarily true, simply because of the meaning of the word "sufficient", as anything that does not in fact bring about a proposed change can be said to be insufficient for doing so. Thus, claims of insufficiency are more like "grammatical remarks" (cf. Wittgenstein, 1953) on the usage of the word "sufficient", rather than substantive suggestions, insofar as they are not expressed in precise quantities (e.g. hours, days, or weeks). Claiming that improvements occur after, say, 16 weeks is a testable claim; claiming that improvements would occur after an indeterminate "longer period" is not. Therefore, future research should aim to specify, in quantifiable terms, the conditions under which cognitive improvements are expected to occur, i.e. what course duration and course intensity is necessary; what achieved level of proficiency in the foreign language is required; and what population the intervention should be aimed at. As long as these conditions are not specified, critics can always resort to counterfactual scenarios when confronted with disconfirming evidence.

On the other hand, it is certainly possible that engaging in more intense language training over a longer period of time would have yielded tangible results, but this raises questions about the feasibility and applicability of such an intervention. If cognition could be improved but only after several months or years of very intense language studies, would it be of any use, as very few (if any) would take part in such an intervention? My guess is "no", as engagement in such activities is costly and more importantly, time-consuming. Any potential cognitive benefits are, at best, a positive side effect of an otherwise stimulating and joyful activity that some people engage in anyhow.

In a broader perspective, the feasibility of broad cognitive improvements through short-term, focal interventions might be questioned. In old age, the cognitive life course trajectory is determined by more than 60 years of ontogenetic experience (in addition to thousands of years of phylogenetic evolution) and it might be somewhat naïve to expect that a simple, three-month intervention would drastically alter that trajectory, lest the intervention entails a much more radical procedure (e.g. surgical interventions; the best way to permanently lost weight is amputation). Even if one might occasionally observe short-term benefits from such interventions, these are likely to be transient.

In a sense, plasticity and stability trade-off against each other: a completely plastic system cannot retain information, while a completely stable system cannot incorporate new knowledge (Mermillod, Bugaiska, & Bonin, 2013). As such, one can speculate if reduced brain plasticity in old age might actually be a neural optimization strategy, albeit with some negative side effects, as with increasing age maintaining and preserving what's left becomes more important than acquiring new skills. This might be reflected in the relative preservation of crystallized abilities (i.e. relying on already acquired knowledge) while fluid abilities, involved in acquiring new skills, decline in aging. Nevertheless, while brain plasticity is reduced in old age, the holy grail of cognitive aging research is to find out how to reverse this process through experimental manipulation.

4.4   Educational attainment and cognitive functioning

It is probably safe to say that going to school has a drastic impact on our cognitive functioning. Education is, in a sense, the ultimate cognitive intervention, lasting for many years and being progressively more challenging, thus giving ample opportunity for people to experience that mismatch between functional supply and environmental demand necessary to induce plastic changes. This thesis adds to the converging evidence that the main benefit of going to school is exerted through its effect on cognitive functioning in early life and young adulthood, i.e. during a period when the brain is more malleable and receptive. There is some evidence to suggest that going additional educational attainment does have causal influence on intelligence (Lager et al., 2016; Ritchie & Tucker-drob, 2018; but see Kremen et al., 2019;). The next question is what the relationship looks like in the lower and upper end of the educational range. It is likely that an extra year at the bottom end, e.g. going from 6 to 7 years, has larger benefit than an extra year at the upper end, e.g. going from 20 to 21 years of education. While the exact nature is yet to be determined, it is not unlikely that educational attainment suffers from diminishing returns.

It is also not known for whom educational attainment is most beneficial. The effect of education likely interacts with a large number of factors, such as parental socio-economic status or innate cognitive ability. One possibility – the compensation scenario – is those with low intelligence will benefit more from additional schooling,

possibly because they have more to gain. Another possibility (perhaps gloomy albeit more plausible given the far-reaching effects of intelligence) is the magnification scenario, whereby those with already high IQ will benefit more from additional schooling. This pattern has been observed for health literacy (Gottfredson & Deary, 2004) in that not only are more intelligent people healthier and live longer in general, but they are also more cost-efficient at consuming medical treatment once they become sick.

Finally, theories of cognitive decline must be updated to incorporate converging evidence that the overall effect of educational attainment on cognitive decline is negligible, and possibly non-existent. There might be 99 factors that are associated with the rate of cognitive decline in old age, but education ain't one.

4.5 Methodological considerations

Over the course of my doctoral education I have found that I tend to gravitate towards methodological rather than substantive issues (sometimes to the dismay and frustration of my supervisors). The following sections do not concern cognitive aging per se, but apply to research more generally – in particular, statistical inference, philosophy of science and the prospect of psychology as a quantitative science.

4.5.1 Frequentist inference

The most common framework for statistical inference in biomedical research is a confused amalgamation of frequentist statistics, i.e. using null hypothesis significance testing (NHST) and p-values. Frequentism is so called because the conception of probability in this framework is defined in terms of long run frequencies, or proportions. Inference is made on the basis of the p-value, which is the probability of observing a test statistic at least as extreme as the obtained test statistic, conditional on the truth of the null hypothesis and other assumptions (Wasserstein & Lazar, 2016). Thus, the p-value is a measure of incompatibility between observed data and a proposed statistical model, the null hypothesis. The test statistic can be thought of as a summary measure, e.g. the ratio of the observed estimate (the "signal") and the associated error (the "noise"). If the signal-to-noise ratio is large enough (e.g. close to 2) the resulting p-value is small enough to warrant rejection of the null hypothesis, implicitly in favor of the alternative hypothesis (which is the hypothesis the researcher typically entertains). In NHST, two kinds of errors can be made: one can falsely declare the presence of an effect which is absent in the population (a "false alarm", or a type 1 error) or one can falsely declare the absence of an effect that is present in the population (a "miss", or a type 2 error). A principal goal of frequentist inference is to put an upper limit on the proportion of type 1 errors, e.g. the proportion of times we incorrectly reject a true null hypothesis. This limit is set by α-level, conventionally at 5%.

Significance testing have been notoriously criticized since their inception due to interpretational difficulties of p-values (Goodman, 2009) and confidence intervals (Hoekstra et al., 2014), that the underlying logic is flawed (Cohen, 1994) and that the foundational assumption of a true null hypothesis always is false (Tukey, 1991). Schmidt & Hunter (1997) called the enterprise "thoroughly discredited" (p. 37) and claimed it "never makes a positive contribution" (p. 62). While proponents of significance testing agree that researchers routinely misinterpret the p-value, they sometimes insist that there is nothing wrong with the p-value per se and that its statistical definition is sound. After all, "guns don't kill people, people do".

P-values are often misinterpreted as the probability that the results are due to chance; the probability that the null hypothesis is true; that the obtained result is a type 1 error; or as the converse probability (1-p) that the results will replicate in a future study (Greenland et al., 2016). Most would agree that the above quantities would be very useful to have, yet they do not correspond to (and have little in common with) the formal definition of the p-value. Those quantities are available, but only for a price, and not everyone is willing to pay that price (see "priors" below). Given that the p-value is confused with the above quantities, it is no wonder that p-values remain popular.

NHST has also been criticized for invalid logical inference (Schmidt & Hunter, 1997). Informally, the logic of NHST is this: if the null hypothesis is true, the observed result is unlikely; but this result did occur; therefore, the null hypothesis is unlikely (Cohen, 1994). While this line of reasoning is suggestive and reminiscent of a valid logical form called modus tollens, it is rendered invalid by the introduction of probabilistic statements (e.g. "unlikely").

Null hypothesis significance testing is sometimes also presented as being falsificationist, in that the aim is to reject a proposed statistical model. Falsificationism (Popper, 1959) is a variant of the problem of induction (Hume, 1748) in that there is an asymmetry between confirmatory and disconfirmatory evidence: confirmatory evidence cannot differentiate between hypotheses, only disconfirming evidence can. Therefore, data inconsistent with predictions are far more informative than data consistent with predictions, so researchers learn more from getting the world wrong than from getting the world right. This is unintuitive, and while many researchers pay lip-service to the notion of falsifiability it is rarely applied in practice.

However, while falsification is a laudable goal, the connection to significance testing is fatally flawed. A fundamental problem is that substantive psychological theories very rarely makes precise numerical predictions. They are, at best, formulated as directional hypotheses (e.g. "A will be larger than B") without specifying how much larger. The only precise prediction is the one made by the null hypothesis (which predicts a difference of 0), which very rarely is of any substantive interest and is typically not the hypothesis the researcher entertains. Furthermore, the term

"hypothesis testing" can be deceiving in that it suggests that the hypothesis being tested is a substantive hypothesis. Let's be clear: the hypothesis being tested is not determined by the mental state of the researcher, i.e. what hypothesis the researcher thinks he or she is testing, but is strictly determined by the statistical test (which is invariably the nil null hypothesis of no difference). This confusion is reflected in formulations like "We tested the hypothesis that… [e.g.] by a paired t-test" and interpreting a significant finding as evidence for the hypothesis. Thus, researchers rarely, if ever, attempt to falsify substantive theories, because substantive theories do not make precise predictions.

One of the most common claims is that null hypothesis significance testing is futile because taken literally, the null hypothesis is always false (Tukey, 1991). While this may seem trivial it critically undermines the entire enterprise of NHST.

First, the null hypothesis is sometimes true simply by construction, e.g. at baseline following proper randomization. One of the main purposes of randomization is to induce independence between treatment condition and other covariates, so that background characteristics cannot interfere with treatment assignment and thereby does not bias the estimator of the average treatment effect (Deaton et al., 2017).

Incidentally, this is also why significance testing of baseline differences in randomized controlled trials (RCTs) is nonsensical. Significance testing of sample means is done to reject the null hypothesis that the means of the populations from which the samples were drawn are equal. However, in an RCT there are no two distinct populations, as everyone is sampled from the same population, and group membership is artificially created through randomization. Thus, rejecting the null hypothesis on the basis of significant baseline differences in an RCT is tantamount to rejecting the necessary truth that the population mean is equal to the population mean, i.e. that x = x. This is clearly absurd (Altman, 1985). Still, p-values routinely accompany baseline between-group comparison in RCTs, and are routinely requested by reviewers. Worse, variables exhibiting significant differences following proper randomization are subsequently included as covariates which leads to unknown error rates.

Second, we may grant that the null hypothesis is true also for certain non-trivial, empirical relationships out there in the world. If we accept Tukey's claim – that no parameter is ever truly zero in the population, or that it is always different from zero at some decimal place – it is not clear why we grant that special status only to the number zero. As an example, consider the following: "people with blue eyes" and "people with brown eyes" refer to fairly well-defined populations. Let H1 and H2 denote the average height (in centimeters) of people in these populations. H1 and H2 are going to be different to some decimal place, for a number of reasons – perhaps substantive, perhaps random – so the difference H1-H2 is

going to be different from 0 to some decimal place. Thus, Tukey would arguably dismiss the null hypothesis H1-H2 = 0 cm a priori. But if a difference of exactly 0 cm can be dismissed, the same can be said of a difference of 2 cm, or 2.1 cm, or 2.159 cm, or any proposed exact parameter value; there's nothing special about the value 0 in that regard. Taken to the extreme, the conclusion must be that the difference H1-H2 does not even denote a well-defined quantity, because the mass probability of any point value is zero.

Third, a direct consequence of Tukey's claim is Paul Meehl's conjecture of the "crud factor" (Meehl, 1990); the assumption that everything affects everything else to some (miniscule) extent. On this view, arguably there is no such thing as "overfitting a model" because every non-zero relationship is to be modeled and estimated. This assumption makes for a very complex and inflated ontology, and ontological parsimony is a virtue that perhaps should not be readily dismissed. There is of course nothing that says that the world cannot be of infinite complexity. But the role of science is not to mirror an infinitely complex world, but to provide useful models of the world, and our scientific models of the world cannot be of infinite complexity if they are to be at all useful.

Fourth, accepting the premise that the null is never literally true also proves fatal to the fundamental purpose of significance testing, which is to put an upper limit on the proportion of type 1 errors (i.e. the erroneous rejection of a true null hypothesis). If we accept that the null hypothesis is never true, then we also have to accept that type 1 errors are impossible because they can only occur when the null hypothesis is true. And it is not clear why we should be concerned about limiting something that never has a chance to occur. Ironically then, we are faced with a paradox: if we agree with Tukey in that the null hypothesis is never literally true, this necessarily implies at least one true null hypothesis, namely the meta-scientific hypothesis that the total number of false positives in the literature is 0%. At least that's good news.

It seems that we may either accept Tukey's claim that there are no true null hypotheses or endorse the ambition of putting an upper limit on the proportion of type 1 errors, but we cannot consistently entertain both of these ideas simultaneously. If we accept the presence of true null hypotheses, ideally we should also employ a consistent method of inference that will arrive at the correct answer in the large sample limit (Rouder, Speckman, Sun, Morey, & Iverson, 2009). Significance testing is inconsistent in that it will converge on the right answer only when the null hypothesis is false, not when it's true. If this (somewhat contrived) reasoning is correct, it seems that significance testing should not be employed regardless of the truth or falsify of the null hypothesis: if the null is true, you will never reach that conclusion using an inconsistent method, and if the null is false, there is no risk of a type 1 error.

Furthermore, finding support for the null hypothesis is pivotal in establishing invariances. Showing that things do not differ across different levels of an independent variable is an important goal in science (Gallistel, 2009), such as proving that effects are additive (i.e the absence of an interaction). In Study III, the substantive hypothesis of passive reserve coincides with the statistical null hypothesis, i.e. that educational attainment is not associated with the rate of cognitive decline. Another case of establishing invariances occurs in structural equation modelling (SEM) when assessing measurement invariance, where the researcher aims to show that imposing path constraints on measurement parameters do not lead to worse fit, as indicated by $p > .05$. Normally one is not allowed to accept the null hypothesis based on $p > .05$, yet within the context of SEM this is somehow deemed legitimate.

Originally, built into the notions of type 1 and type 2 errors was the idea that one should balance the two different kinds of errors, or weight their relative cost against each other (Neyman & Pearson, 1933). In practice, however, every researcher sets the acceptable type 1 error rate to 5%. For an individual researcher the cost of making a type 1 error is arguably almost non-existent, while the potential benefit of making a type 1 error includes a potential publication in a high impact journal. Conversely, the cost of a type 2 error (e.g. failure to detect a true effect) is much higher, as it may indicate sloppiness on part of the researcher (e.g. lack of "flair"; Baumeister, 2016) while conferring no obvious benefit. This might contribute to the alleged large proportion of type 1 errors in the literature (e.g. Ioannidis, 2005).

But the fundamental problem with p-values is ultimately a practical matter in how they guide data-analytic decision making. Scientists have been conditioned to obtain statistical significance, and the first thing we do when looking at data analysis output is to seek asterisks in the statistical significance column ("star-gazing"; cf. McElreath, 2016). A near-significant result of p = .071 bothers us much more than p = .71 in that we try to transmute the former into a significant result through creative analysis and post-hoc reasoning, while the latter make us look elsewhere. We also rarely know the functional form of a relationship between two variables with arbitrary scaling, so who's to say that the underlying relationship is strictly linear rather than a complex polynomial, or that the cut-off for an outlier should be 2 rather than 1.5 standard deviations? A researcher faces many choices during data analysis, leading to a multitude of "researcher degrees of freedom" (Simmons, Nelson, & Simonsohn, 2011). When "exploring the data set" (a euphemism for trying out various analytical methods) obtaining statistical significance for one analytic choice over another can itself be interpreted as validation that one is on the right track: e.g. "using this transformation, or this factor analytic rotation, or including this covariate, or excluding these participants, yields statistical significance, which means I'm picking up a signal, so there's something there".

Data-analytic decision making on the basis of p-values has been labeled "p-hacking" (Simmons et al., 2011) or analogous to a walk down the "garden of forking paths" (Gelman & Loken, 2013). I think the latter is preferable because "p-hacking" is an

active verb with connotations implying something sinister on part of the researcher, whereas the metaphor of the garden of forking suggest blissful ignorance rather than malicious intent (c.f. Hanlon's razor). But like the road to hell, the garden of forking paths is paved with good intentions.

The problem is that once we find what we're looking for, then we stop looking. While this makes sense in everyday situations, e.g. when looking for one's misplaced keys, in significance testing it undermines the interpretation of the p-value. The definition of the p-value is based on counter-factual and hypothetical long-run frequencies of what would happen if we were to repeat the same experiment (and analysis) a very large number of times. Given that data-analytic choices depend on the outcome of certain statistical tests, it is not even clear what "the same experiment" means, because the sampling space is not uniquely determined. In the next data analysis of an identical experiment, different data-analytical choices would be made due to idiosyncratic properties of the data set.

The prospect of pre-registration (of planned data collection and data analysis) has received wide spread attention as a possible remedy to such data-analytic flexibility. Pre-registrations can also help distinguishing between confirmatory and exploratory research. It can also help with yielding valid p-value, e.g. by fixating the reference for what "the same experiment" means. However, pre-registration can only help insofar as they actually restrict researcher degrees of freedom and limit data-analytic flexibility, and not all pre-registrations are created equal; there is currently large variability in the quality and scope of pre-registrations (Veldkamp et al., 2019). On the one end of the spectrum are pre-registrations specifying how data collection, data cleaning and data analysis is to be done once the data is in, down to the level of supplying pre-specified analysis code, along with precise numerical hypotheses. On the other end of the spectrum are pre-registrations stating directional predictions and nothing else. At worst, insufficiently precise pre-registrations allow the researcher to pass off the research as being pre-registered, with all the potential benefits of doing so, while keeping the actual pre-registration vague enough so as to still allow substantial analytic flexibility. An extreme position, though not entirely unreasonable, would be that all experiments that are not pre-registered (i.e. almost all research conducted prior to 2012 when pre-registration took off) should be deemed exploratory. There is of course nothing wrong with exploratory research, as long as it is presented as such. Indeed, given the lack of precision in psychological hypotheses, perhaps we would be better off conducting only exploratory research until we come up with a theory that's actually testable.

4.5.2   Bayesian inference

My disenchantment with frequentist statistics occupied much of my PhD and led me towards the most common alternative to frequentist inference, namely Bayesian statistics. One fundamental difference between these two inferential paradigms

lies in the concept of probability: while frequentists conceive of probability as a limiting proportion, e.g. a long run frequency, Bayesians interpret probability as quantifying degrees of belief. This latter probability is represented by the introduction of a prior distribution.

Another difference between the two is that frequentists calculate the probability of observed data conditional on a given parameter value ("the model is fixed, the data is random") while Bayesians calculate the probability of different parameter values, conditional on observed data ("the data is fixed, the model is random"). The prior distribution of the parameter is combined with the likelihood from the data, resulting in a posterior distribution of the parameter.

Whereas frequentist inference use p-values for selection, and point estimates and confidence intervals for estimation, Bayesian statistics use Bayes factors for model selection, and posterior means and highest density intervals for estimation. Thus, Bayesian inference can accommodate probability statements pertaining to hypotheses and parameter estimates, such as "the probability that the null hypothesis is true" and "there is a 95% probability that the true value lies within the confidence limits". Such statements are not possible within the frequentist framework, as parameters do not have probabilities associated with them (a hypothesis is either true or false, and a parameter value either lies within the confidence limit or it doesn't – it is not a matter of probability to a frequentist).

Yet another difference is that within the frequentist framework, parameter estimation and hypothesis testing are two sides of the same coin (e.g. the null hypothesis is rejected if the estimate is about twice as large as the standard error). In the Bayesian framework parameter estimation and hypothesis testing are quite different, in particular with regards to the role of the prior. In parameter estimation, the influence of the prior on the resulting posterior is quickly diminished as more data accumulates ("the swamping of the prior"). In hypothesis testing, such "swamping" doesn't occur as the role of the prior is to represent a formalization of the alternative hypothesis, and thus is never rendered obsolete. This makes the specification of the prior distribution pivotal in Bayesian hypothesis testing, while it is much less important in parameter estimation (Lee & Wagenmakers, 2014).

Whereas the p-value is infamous for being difficult to interpret, Bayes factors are fairly straightforward: they are the ratio of the probability of the data under two competing models. A Bayes factor of 5 simply means that the observed data is 5 times more likely under one model than under a competing model. One advantage of Bayesian hypothesis testing is the ability to quantify evidence in favor of the null hypothesis, seen in Study III (and somewhat in Study II). Employed within the context of null hypothesis testing, the Bayes factor can be easily approximated using the Savage-Dickey density ratio (Lee & Wagenmakers, 2014) by dividing the ratio of the prior and the posterior at the parameter value of interest, e.g. 0.

As such, Bayes factors are marginal measures of evidence suggesting how one should update one's belief after seeing new data. An example of this is the findings in Study III, suggesting Bayes factors close to 1 meaning that one should not change one's belief at all (given that one agrees with the prior specification). We also show that if one accepts a vaguer prior (corresponding to larger effects on the role of educational attainment on the rate of episodic memory decline) there is increasing evidence against the alternative hypothesis.

One can utilize Bayesian statistics for many reasons. One reason is computational tractability. Some hierarchical models involving a large number of parameters and few observations for each level, and are so complex that they cannot easily be estimated using frequentist methods. In contrast, Bayesian implementations of such models might be straight forward due to the introduction of regularizing priors. Another advantage is that the prior distribution allows the Bayesian to explicitly bring external information, i.e. previous research, to the model, thus formalizing the cumulative nature of science. A frequentist, on the other hand, leaves previous research at the door as soon as he begins data analysis, in that he always starts out from a point of ignorance, only informed by the current data set. This ability to formally "stand on the shoulders of giants" is especially beneficial in small sample research, when gathering data is expensive or samples are rare (e.g. in studying rare diseases).

This thesis consists of neither overly complex models, nor particularly small sample research. In both articles where Bayesian methods have been employed (Study II and III) the same models can easily be estimated using frequentist methods, and would probably have arrived at similar conclusions. The posterior distributions are informed mainly by the likelihood from the available data, using uninformative priors for all parameters except those subject to hypothesis testing.

Instead, it was my interest in the philosophy of science, along with a very open-minded main supervisor, that prompted the use of Bayesian statistics in studies II and III. The excursion to learn an alternative statistical paradigm has not been altogether productive and has arguably taken an excessive amount of time from substantive research (which is mirrored in the present thesis). My own experience is that the vast majority of researchers who employ frequentist methods do so not because of a reasoned theoretical argument in favor of significance testing (incidentally, I have yet to see one) but because of tradition and pragmatics, e.g. that's what introductory statistics courses teach. This might also explain why calls to abandon significance testing has fallen on deaf ears, as it is difficult to reason someone out of a position that they did not reason themselves into in the first place. Yet another, perhaps bigger, practical obstacle for the dissemination of Bayesian statistics is the obnoxious missionary zeal and smugness newly-convert Bayesians tend to flaunt. And with friends like that you don't need enemies.

5   ACKNOWLEDGEMENTS

6   REFERENCES

Aichele, S., Rabbitt, P., & Ghisletta, P. (2015). Life Span Decrements in Fluid Intelligence and Processing Speed Predict Mortality Risk. *Psychology and Aging*, *30*(3), 598–612.

Altman, D. G. (1985). Comparability of Randomised Groups. *The Statistician*, *34*(1), 125. http://doi.org/10.2307/2987510

Antoniou, M., Gunasekera, G. M., & Wong, P. C. M. (2013). Foreign language training as cognitive therapy for age-related cognitive decline: A hypothesis for future research. *Neuroscience and Biobehavioral Reviews*, *37*(10), 2689–2698. http://doi.org/10.1016/j.neubiorev.2013.09.004

Baddeley, A. (1992). Working Memory. *Science*, *255*, 556–559.

Baltes, P. B. (1968). Longitudinal and Cross-Sectional Sequences in the Study of Age and Generation Effects. *Human Development*, *11*(3), 145–171.

Baumeister, R. F. (2016). Charting the future of social psychology on stormy seas: Winners, losers, and recommendations. *Journal of Experimental Social Psychology*, *66*, 153–158. http://doi.org/10.1016/j.jesp.2016.02.003

Becker, N., Kalpouzos, G., & Salami, A. (n.d.). Structure-function associations of successful associative encoding.

Becker, N., Laukka, E. J., Kalpouzos, G., Naveh-Benjamin, M., Bäckman, L., & Brehmer, Y. (2015). Structural brain correlates of associative memory in older adults. *NeuroImage*, *118*, 146–153. http://doi.org/10.1016/j.neuroimage.2015.06.002

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism : consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–250. http://doi.org/10.1016/j.tics.2012.03.001

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A., Sommer, J., Wolbers, T., & Knecht, S. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage*, *25*, 958–968. http://doi.org/10.1016/j.neuroimage.2004.12.019

Burgess, N., Maguire, E. A., & Keefe, J. O. (2002). The Human Hippocampus and Spatial and Episodic Memory. *Neuron*, *35*, 625–641.

Cattell, R. B. (1963). THEORY OF FLUID AND CRYSTALLIZED INTELLIGENCE : *Journal of Educational Psychology*, *54*(1), 1–22.

Ceci, S. J. (1991). How Much Does Schooling Influence General Intelligence and Its Cognitive Components ? A Reassessment of the Evidence. *Developmental Psychology*, *27*(5), 703–722.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, 49(12), 997-1003. http://dx.doi.org/10.1037/0003-066X.49.12.997

Colom, R., Lluis-font, J. M., & Andre, A. (2005). The generational intelligence gains are caused by decreasing variance in the lower half of the distribution : Supporting evidence for the nutrition hypothesis. *Intelligence*, *33*, 83–91. http://doi.org/10.1016/j.intell.2004.07.010

Davachi, L. (2006). Item , context and relational episodic encoding in humans. *Curr Opin Neurobiol*, *16*, 693–700. http://doi.org/10.1016/j.conb.2006.10.012

Deaton, A., Cartwright, N. (2018) Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, *210*(8), 2-21.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: changes in grey matter induced by training. *Nature*, *427*, 311–312.

Finkel, D., Andel, R., Gatz, M., & Pedersen, N. L. (2009). The Role of Occupational Complexity in Trajectories of Cognitive Aging Before and After Retirement, *Psychology and Aging, 24*(3), 563–573. http://doi.org/10.1037/a0015511

Flynn, J. R. (1984). The Mean IQ of Americans : Massive Gains 1932 to 1978. *Psychological Bulletin*, *95*(1), 29–51.

Gallistel, C. R. (2009). The Importance of Proving the Null. *Psychological Review*, *116*(2), 439–453. http://doi.org/10.1037/a0015251

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Gerstorf, D., Ram, N., Hoppmann, C., Willis, S. L., & Schaie, K. W. (2011). Cohort Differences in Cognitive Aging and Terminal Decline in the Seattle Longitudinal Study. *Developmental Psychology*, *47*(4), 1026–1041. http://doi.org/10.1037/a0023426

Goodman, S. (2009). A dirty dozen: 12 p-value misconceptions. Semin Hematol, 45(3):135-140. doi: 10.1053/j.seminhematol.2008.04.003.

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence Predicts Health and Longevity, but Why ? *Current Directions in Psychological Science*, *13*(1), 1–4.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337–350. http://doi.org/10.1007/s10654-016-0149-3

Hume, D., & Millican, P. F. (1748/2007). An enquiry concerning human understanding. Oxford: Oxford University Press.

Ioannidis, JPA (2005). Why Most Published Research Findings Are False. *PLoS Med*, 2(8): e124: https://doi.org/10.1371/journal.pmed.0020124

Karbach, J., & Verhaeghen, P. (2014). Making Working Memory Work : A Meta-Analysis of Executive-Control and Working Memory Training in Older Adults. *Psychological Science*, *25*(11), 2027–2037. http://doi.org/10.1177/0956797614548725

Kempermann, G., Song, H., & Gage, F. H. (2015). Neurogenesis in the Adult Hippocampus. *Cold Spring Harb Perspect Biol.*, *7*(9).

Köhncke, Y., Laukka, E. J., Brehmer, Y., Kalpouzos, G., Li, T., Fratiglioni, L., … Lövdén, M. (2016). Neurobiology of Aging Three-year changes in leisure activities are associated with concurrent changes in white matter microstructure and perceptual speed in individuals aged 80 years and older. *Neurobiology of Aging*, *41*, 173–186. http://doi.org/10.1016/j.neurobiolaging.2016.02.013

Kremen, W. S., Beck, A., Elman, J. A., Gustavson, D. E., Reynolds, C. A., & Tu, X. M. (2019). Influence of young adult cognitive ability and additional education on later-life cognition. *PNAS*, *116*(6), 2021–2026. http://doi.org/10.1073/pnas.1811537116

Lager, A., Seblova, D., & Falkstedt, D. (2016). Cognitive and emotional outcomes after prolonged education : a quasi-experiment on 320 182 Swedish boys. *International Journal of Epidemiology*, *46*(1), 303–311. http://doi.org/10.1093/ije/dyw093

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian Cognitive Modeling: A Practical Course.* http://doi.org/10.1017/CBO9781139087759

Lenehan, M. E., Summers, M. J., Saunders, N. L., Summers, J. J., & Vickers, J. C. (2014). Relationship between education and age-related cognitive decline : a review of recent research. *Psychogeriatrics*, *15*(2), 154–162. http://doi.org/10.1111/psyg.12083

Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010). A Theoretical Framework for the Study of Adult Cognitive Plasticity. *Psychological Bulletin*, *136*(4), 659–676. http://doi.org/10.1037/a0020080

Lynn, R. (1990). THE ROLE OF NUTRITION IN SECULAR INCREASES IN INTELLIGENCE. *Person. Individ. Diff*, *11*(3), 273–285.

Mårtensson, J., Eriksson, J., Christian, N., Lindgren, M., Johansson, M., Nyberg, L., & Lövdén, M. (2012). Growth of language-related brain areas after foreign language learning. *NeuroImage*, *63*(1), 240–244. http://doi.org/10.1016/j.neuroimage.2012.06.043

McElreath, R. (2016). rethinking: an R Package for Fitting and Manipulating Bayesian Models, (1), 1–26.

Meehl, P. E. (1990). Why Summaries of Research on Psychological Theories Are Often Uninterpretable. *Psychological Reports*, *66*(1), 195. http://doi.org/10.2466/pr0.1990.66.1.195

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–291. http://doi.org/10.1037/a0028228

Mermillod, M., Bugaiska, A., & Bonin, P. (2013). The stability-plasticity dilemma : investigating the continuum from catastrophic forgetting to age-limited learning effects. *Frontiers in Psychology*, *4*(August), 1–3. http://doi.org/10.3389/fpsyg.2013.00504

Naveh-Benjamin, M. (2000). Adult Age Differences in Memory Performance : Tests of an Associative Deficit Hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1170–1187. http://doi.org/10.1037/0278-7393.26.5.1170

Neisser, U. (1997). Rising Scores on Intelligence Tests: Test scores are certainly going up all over the world, but whether intelligence itself has risen remains controversial. *American Scientist*, 85(5), 440-447.

Neyman, J., & Pearson, E. S. (1933). On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philosophical Transactions Of the Royal Society Of London. Series A, Containing Papers Of a Mathematical or Physical Character*, *231*, 289–337.

Nilsson, L., Bäckman, L., Erngrund, K., Adolfsson, R., Bucht, G., & Karlsson, S. (1997). The Betula Prospective Cohort Study : Memory, Health, and Aging. *Aging, Neuropsychology, and Cognition*, *4*(1), 1–32. http://doi.org/10.1080/13825589708256633

Opdebeeck, C., Martyr, A., & Clare, L. (2015). Cognitive reserve and cognitive function in healthy older people : a meta-analysis. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn.*, *23*(1), 40–60. http://doi.org/10.1080/13825585.2015.1041450

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *CORTEX*, *69*, 265–278. http://doi.org/10.1016/j.cortex.2015.04.014

Penfield, W., & Milner, B. (1958). Memory Deficit Produced by Bilateral Lesions in the Hippocampal Zone. *A.M.A. Archives of Neurology and Psychiatry*, *79*, 475–497.

Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., … Acker, J. D. (2005). Regional Brain Changes in Aging Healthy Adults : General Trends , Individual Differences and Modifiers. *Cerebral Cortex*, *15*(November), 1676–1689. http://doi.org/10.1093/cercor/bhi044

Ritchie, S. J., & Tucker-drob, E. M. (2018). How Much Does Education Improve Intelligence ? A Meta-Analysis. *Psychological Science*, *29*(8). http://doi.org/10.1177/0956797618774253

Ritchie, S. J., Tucker-drob, E. M., Cox, S. R., Corley, J., Dykiert, D., Redmond, P., … Deary, I. J. (2016). Intelligence Predictors of ageing-related decline across multiple cognitive functions. *Intelligence*, *59*, 115–126. http://doi.org/10.1016/j.intell.2016.08.007

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, *35*, 83–92. http://doi.org/10.1016/j.intell.2006.05.004

Rönnlund, M., & Nilsson, L. G. (2009). Flynn effects on sub-factors of episodic and semantic memory: Parallel gains over time and the same set of determining factors. *Neuropsychologia*, *47*(11), 2174–2180. http://doi.org/10.1016/j.neuropsychologia.2008.11.007

Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L.-G. (2005). Stability, Growth, and Decline in Adult Life Span Development of Declarative Memory: Cross-Sectional and Longitudinal Data From a Population-Based Study. *Psychology and Aging*, *20*(1), 3–18. http://doi.org/10.1037/0882-7974.20.1.3

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, *16*(2), 225–237. http://doi.org/10.3758/PBR.16.2.225

Salthouse, T. A. (2009). When does age-related cognitive decline begin ? *Neurobiology of Aging*, *30*, 507–514. http://doi.org/10.1016/j.neurobiolaging.2008.09.023

Schaie, K. W. (1965). A GENERAL MODEL FOR THE STUDY OF DEVELOPMENTAL PROBLEMS. *Psychological Bulletin*, *64*(2), 92–107.

Schaie, K. W. (1994). The Course of Adult Intellectual Development. *American Psychologist*, *49*(4), 304–313.

Schaie, K. W. (2005). What Can We Learn From Longitudinal Studies of Adult Development? *Research in Human Development*, *2*(3), 133–158. http://doi.org/10.1207/s15427617rhd0203_4

Schaie, K. W. (2007). Generational differences; the age-cohort-period model. In: Birren JE, editor. *Encyclopedia of gerontology: Age, aging, and the aged*. 2nd edit. Oxford, UK: Elsevier; pp. 601–609.

Schmidt, F. L., & Hunter, J. E. (1997). Eight Common But False Objections to the Discontinuation of Significance Testing in the Analysis of Research Data. In Harlow, LL., Mulaik, S., Steiger, H. (Eds). *What if there were no significance tests* (pp. 37–64).

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology : Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366. http://doi.org/10.1177/0956797611417632

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do "Brain-Training" Programs Work? *Psychological Science in the Public Interest*, *17*(3), 103–186. http://doi.org/10.1177/1529100616661983

Stern, Y. (2002). What is cognitive reserve? Theory and research applicaion of the reserve concept. *Journal of the International Nueropsychological Society*, *8*(2002), 448–460. http://doi.org/10.1017.S1355617701020240

Stern, Y. (2010). Cognitive reserve. *Neuropsychologia*, *47*(10), 2015–2028. http://doi.org/10.1016/j.neuropsychologia.2009.03.004.Cognitive

Strenze, T. (2007). Intelligence and socioeconomic success : A meta-analytic review of longitudinal research ☆. *Intelligence*, *35*, 401–426. http://doi.org/10.1016/j.intell.2006.09.004

Sundet, J. M., Eriksen, W., & Tambs, K. (2008). Intelligence Correlations Between Brothers Decrease With Increasing Age Difference Evidence for Shared Environmental Effects in Young Adults. *Psychological Science*, *19*(9), 843–847.

Tucker-Drob, E. M. (2011). Global and Domain-Specific Changes in Cognition Throughout Adulthood. *Developmental Psychology*, *47*(2), 331–343. http://doi.org/10.1037/a0021361

Tukey, J. W. (1991). The Philosophy of Multiple Comparisons. *Statistical Science*, *6*(1), 100–116.

Tulving, E. (1972). Episodic and semantic memory. In Tulving, E., Donaldson, W. (Eds.), *Organization of Memory* (pp. 381-403). New York: Academic Press.

Veldkamp, C. L. S., Bakker, M., van Assen, M. A. L. M., Crompvoets, E. A. V., Ong, H. H., Nosek, B. A., … Wicherts, J. M. (2018, September 4). Ensuring the quality and specificity of preregistrations. https://doi.org/10.31234/osf.io/cdgyh

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's Statement on p-Values: Context, Process, and Purpose. *American Statistician*, *70*(2), 129–133. http://doi.org/10.1080/00031305.2016.1154108

Wenger, E., Brozzoli, C., Lindenberger, U., & Lövdén, M. (2017). Expansion and Renormalization of Human Brain Structure During Skill Acquisition. *Trends in Cognitive Sciences*, *21*(12), 930–939. http://doi.org/10.1016/j.tics.2017.09.008

Wenger, E., Kühn, S., Verrel, J., Mårtensson, J., Bodammer, N. C., Lindenberger, U., & Lövdén, M. (2016). Repeated Structural Imaging Reveals Nonlinear Progression of Experience-Dependent Volume Changes in Human Motor Cortex. *Cerebral Cortex*, *27*(5), 1–15. http://doi.org/10.1093/cercor/bhw141

Wittgenstein, L. (1953) Philosohipcal Investigations. New York, USA: Macmillan.

Zahodne, L. B., Glymour, M. M., Sparks, C., Bontempo, D., Dixon, R. A., MacDonald, S. W. S., & Manly, J. J. (2011). Education does not slow cognitive decline with aging: 12-year evidence from the victoria longitudinal study. *Journal of the International Neuropsychological Society*, *17*(6), 1039–1046. http://doi.org/10.1017/S1355617711001044

Karolinska
Institutet