

4. Method

The Method section first provides an overview of the Intra-person Dynamics study of which this dissertation work forms part. Sample descriptions and procedural details are reported next. Special care has been taken to explain data-handling procedures, since these issues are rather complex in the context of research on intraindividual variability. Specifically, the Method chapter contains some sections in which general issues concerning the empirical investigation of intraindividual variability are linked to the specific approaches taken in this dissertation. Finally, the steps of analysis reported in the Results section and the dependent variables used in these analyses are outlined.

4.1 Overview of the Intra-Person Dynamics Study

The data used for this dissertation work is part of a larger data set obtained in the context of the Intra-Person Dynamics study initiated by the Center for Lifespan Psychology at the Max Planck Institute for Human Development (S.-C. Li et al., 2005). The principal investigators of the study are Shu-Chen Li, Ulman Lindenberger, Jacqui Smith, and Paul B. Baltes. The Intra-Person Dynamics study pursued three main goals. The first was to attain comprehensive, initial descriptions of intraindividual variability in four key domains of psychological functioning, namely well-being (e.g., negative and positive emotions), self-regulation (e.g., motivation and subjective confidence), cognition (e.g., working memory and perceptual speed) and sensorimotor performance (postural control), on various time scales. Second, the study was designed to investigate aging-related differences in intraindividual variability across different domains of psychological functioning and across a variety of observed measures. The third goal was to examine couplings across domains of functioning at the level of individuals and to specify whether aging-related differences in the strength of these couplings can be found. Thus, the Intra-Person Dynamics study was specifically designed to extend prior research on age differences in interrelations between different domains of psychological functioning, which has been based so far mostly on cross-sectional and between-person longitudinal analyses, to the within-person level.

To achieve these goals, the Intra-Person Dynamics Study used a micro-longitudinal study design; in the sense that the participants were assessed very frequently within the duration of the study. For the Intra-Person Dynamics study, it was decided that young and older participants should be repeatedly assessed both within-sessions and over a 45-day period. The data obtained is, therefore, suited for within-person and between-group comparisons with respect to day-to-day

performance fluctuations and within-person couplings of day-to-day performance fluctuations in different psychological functions. Figure 3 displays the overall design of the study.

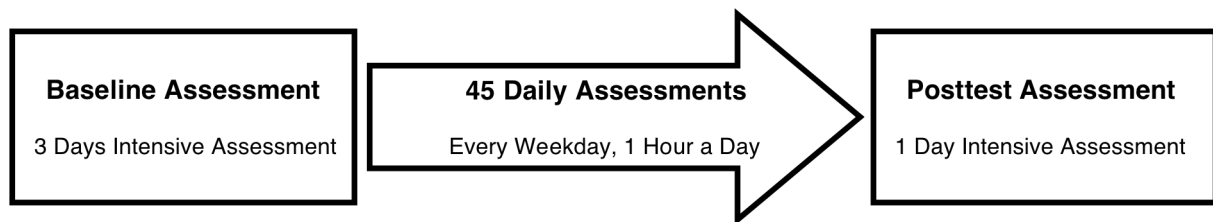


Figure 3. *Design of the Intra-Person Dynamics Study.*

The Intra-Person Dynamics study design consisted of three parts: a comprehensive baseline assessment, a daily assessment period, and a post-test assessment. At the baseline assessment, participants completed a comprehensive assessment battery on three consecutive days before entering the daily assessment period. During the daily assessment period, the participants were tested in approximately 45 daily sessions that took place every weekday (i.e., from Monday to Friday) and lasted one hour. In the post-test assessments, a subset of tests measured at baseline assessment were repeated over one day to evaluate training and transfer effects. All assessments took place at the Max Planck Institute for Human Development. This dissertation focuses exclusively on the variables assessed at the baseline assessment and over the daily assessment period. Therefore, only the procedural details of the baseline and daily assessments are detailed in the remainder of this section.

4.2 Sample

The recruitment procedure, a description of the exclusion criteria for participants, and the details of the assessment periods are all described in this section. In the last subsection of the sample description, the sociodemographic characteristics of the effective sample used in the analyses are presented. The selectivity of the sample in terms of intellectual functioning is evaluated by comparing it to a more representative cross-sectional sample.

4.2.1 Recruitment Procedure and Assessment Periods

The participants were recruited in a number of different ways. Some were recruited via newspaper and radio advertisements, while others joined the study due to recommendations from participants already in the study. Other participants were drawn from the pool of participants

used by the Center of Lifespan Psychology of the Max Planck Institute for Human Development and contacted directly.

Before attending the study, participants were contacted via telephone by an experienced interviewer. In the telephone interview, the interviewers informed potential participants about the practical aspects of the study (i.e. the length of the assessment periods, the duration of the single assessments, and remuneration). During the telephone interview, participants were also asked if they were able to stand unaided for a period of about 10 min without any problem. Only people who responded positively to this question were invited to the institute for further assessments. All participants were informed beforehand that the study was intended to assess cognitive and sensorimotor performances and emotional states in every-day life and that the study included an intensive daily assessment period. However, subjects were not informed about the hypotheses guiding this study.

4.2.2 Assessment Periods

For practical reasons (i.e. limited space and the limited number of people trained to carry out the assessment), the assessment of the complete sample occurred in three separate assessment periods. The first assessment wave took place from October to November 2003 and included nine young adults and seven older adults. The second measurement period started in January 2004 and lasted until April of the same year. The sample of participants assessed over the second daily period consisted of nine young and eleven older adults. After preliminary analyses of the data from the first two assessment periods, it became clear that some participants were not suitable for inclusion in the analyses. Therefore, a third period of assessment was initiated to make up for this. The sample of the third assessment period, which lasted from June to August 2004, consisted of two young adults and five older adults. Participants were paid 10 Euro per hour of assessment. The participants received an additional bonus of 200 Euros for successfully completing the daily assessment period, which was defined as providing data in at least 40 daily sessions out of 45.

4.2.3 Exclusion Criteria for Participants

Initially, 45 adults altogether were recruited for daily assessments in the context of the dynamics study. Of these 45 participants, nine were not included in the analyses reported in this dissertation work for the following reasons: One young and one older woman only actually attended the baseline assessment. The older woman did not take part in the daily assessments because of her lack of German language skills. The young woman was excluded from daily

assessment because she had received medical treatment for a mental illness. Another young woman and another older woman could not be included in the analyses because of a low participation rate in the daily assessments. This was due to the fact that the young woman accepted a job offer, while the older woman fell and broke her arm during the assessment period - this occurred outside the laboratory on a weekend when no assessment was scheduled. Three of the older women were not able to understand the spatial working memory task. All three women also seemed to be lacking in motivation with regards to completing the task properly. Including these three women in the sample would have undermined the validity of the present analysis. One older man was excluded from the analyses because he had a low daily participation rate and also had problems understanding the working memory tasks throughout the assessment period. Finally, one young woman could not be included in the effective sample because of repeated drug use during the assessment period.

At the first glance, the percentage of participants excluded from the analyses appears to be quite high. However, this exclusion rate needs to be put into the context of a micro-longitudinal design of the study. First, the exclusion of participants who do not provide enough repeated measurement data is a unique feature of micro-longitudinal assessments. Second, the micro-longitudinal, repeated measurement design enabled the investigators to detect issues (e.g., drug abuse or a general lack of motivation regarding the working memory task) that might have gone unnoticed in standard cross-sectional or longitudinal designs.

4.2.4 Sociodemographic Characteristics of the Sample

The effective micro-longitudinal sample used in the analyses consisted of 18 young adults and 18 older adults evenly distributed by gender. The mean age of the 18 young adults was 25.5 years and the mean age of the 18 individuals in the older adult group was 74.19 years. The socio-demographic characteristics recorded included chronological age, marital status, the level of education (i.e., the type of German school-leaving qualification), the number of years of formal schooling, subjective health, and the current occupation. All variables were assessed at the baseline assessment. Table 1 displays sociodemographic information about the sample.

Table 1. *Sociodemographic Sample Characteristics at the Baseline Assessment*

	Young adults (N = 18)		Older adults (N = 18)	
Age (years)	Mean	SD	Mean	SD
	25.50	2.73	74.19	2.84
Marital status	N	%	N	%
Unmarried	15	83.3	1	5.6
Long-term Relationship	3	16.7	0	0
Married	0	0	10	55.6
Divorced	0	0	4	22.2
Widowed	0	0	3	16.7
Education	N	%	N	%
Elementary school	0	0	3	16.7
Secondary school level	1	5.6	0	0
High school (12 th / 13 th grade)	16	88.9	7	38.9
College / University	1	5.6	7	38.9
Other	0	0	1	5.6
Years of formal schooling	Mean	SD	Mean	SD
	15.53	3.59	12.67	4.42
Current occupation	N	%	N	%
University student	14	77.8	0	0
Vocational trainee	1	5.6	0	0
Unemployed	3	16.7	0	0
Retired	0	0	18	100

The sample of young participants contained primarily university students, whereas all individuals in the sample of older participants were retired at the time of assessment. The heterogeneity with regard to the level of formal education was much larger in the sample of older adults than in the sample of young adults. This finding is presumably a reflection of the Second World War, which had a large impact on educational opportunities for older individuals. On average, the sample of young adults had received more years of formal schooling than the sample of older adults, $F(1.34) = 4.54$, $p < .05$, $\eta_p^2 = .12$ (The sample heterogeneity did not differ between age groups, Levene's test of equality of error variances $F(1.34) = 1.66$, $p > .10$).

Subjective health was measured with a single item Likert scale rating. The scale ranged from one to five. A scale value of one indicated a poor subjective health. On average, the young adults subjectively rated their health better than the sample of older adults, $F(1.34) = 14.58$,

$p < .05$, $\eta_p^2 = .30$ (sample heterogeneity did not differ between age groups, Levene's test of equality of error variances $F(1.34) = 0.99$, $p > .10$).

4.2.5 Sample Characteristics in Intellectual Functioning

Intellectual functioning was assessed using four marker tests, which can be categorized with respect to the distinction between fluid and crystallized intellectual abilities (Cattell, 1971; Horn & Cattell, 1966) or the decomposition of cognitive functioning into the mechanics and the pragmatics of cognition that originated from the perspective of lifespan development (Baltes, 1987; Baltes et al., 1998). The four marker psychometric tests were Digit Symbol Substitution, Identical Pictures, Spot-a-Word, and Vocabulary. These particular tests were a subsample of the intelligence tests used in the psychometric battery of the Berlin Aging Study (see Lindenberger, Mayr, & Kliegl, 1993, for details on the four tests). Factorial analyses with different samples of young and older adults have shown that the first two psychometric tests can be thought of as a marker test for the mechanics of cognition, whereas the latter two tests may serve as marker tests for the pragmatics of cognition (Baltes & Lindenberger, 1997; S.-C. Li et al., 2004; Lindenberger & Baltes, 1994). Table 2 provides age-comparative sample descriptions with regard to intellectual functioning in these four tests.

Table 2. *Intellectual Functioning in Four Psychometric Tests in Both Age Groups*

	Young Sample (N = 18)		Older Sample (N = 18)	
Cognitive Mechanics	Mean	SD	Mean	SD
Digit Symbol Substitution	59.72	11.83	36.72	11.16
Identical Pictures ^{a)}	42.71	2.93	27.22	3.47
Cognitive Pragmatics	Mean	SD	Mean	SD
Spot-a-Word	21.06	5.86	28.11	3.61
Vocabulary	24.28	3.82	24.72	5.59

^{a)} *The Identical Pictures data from 1 young adult was missing. The effective sample size of the young adult group in the Identical Pictures Test is 17.*

Intellectual functioning was measured using different scales for each of the four tests. The size of the complete sample of 36 participants did not seem to justify the standardization of the individual test scores into a common metric. The age differences in the test scores were, therefore, analyzed with four separate Univariate ANOVAs for each test. In line with the prediction made from a lifespan perspective on cognitive development, the sample of young adults had higher scores than the sample of older adults in both of the tests that indicate cognitive mechanics. The age difference was statistically significant in the Digit Symbol Substitution test, $F(1.34) = 35.99$, $p < .05$, $\eta_p^2 = .51$, and also significant for the Identical Pictures

test, $F(1.33) = 201.87, p < .05, \eta_p^2 = .86$. In line with the prediction of the lifespan perspective on cognitive development, the picture was different in the two tests of cognitive pragmatics. The sample of older adults scored equally high on the Vocabulary test as the young adults, $F(1.34) = 0.08, p > .10, \eta_p^2 = .00$, and even higher than the sample of younger adults in the Spot-a-Word test, $F(1.34) = 18.93, p < .05, \eta_p^2 = .36$. These findings are in accordance with the research regarding age differences in word knowledge. Verhaeghen (2003) demonstrated in a meta-analysis of 324 independent age contrasts that young adults and older adults do not differ cross-sectionally in terms of vocabulary tests involving production but that older adults have a slight advantage over young adults in tests of world knowledge involving recognition.

4.2.6 Examining Sample Selectivity in Intellectual Functioning

Given the particular micro-longitudinal, high intensive measurement design of the Intra-Person Dynamics study, it is important to check the degree to which the sample used in the analyses is similar to samples conventionally used in age-comparative studies. To address this issue, the sample of young adults and the sample of older adults analyzed in this dissertation were contrasted in terms of their intellectual functioning with a young sample and an older sample from a large-scale, representative study. The comparison samples were drawn from the CoOP Mind study conducted by the Max Planck Institute for Human Development in Berlin and the Max Planck Institute for Psychological Research (see S.-C. Li et al., 2004). A particular design feature of the CoOP Mind study is that its participants were randomly drawn from the Berlin City Registry and are, therefore, more representative than a convenience sample (S.-C. Li et al., 2004). The comparison sample contained 24 young adults and 30 older adults. Univariate ANOVAs conducted separately for young adults and older adults demonstrated that both age groups drawn from the COP Mind study matched their counterparts from the Intra Person Dynamics study in terms of the mean age (young adults: $F(1.40) = 0.16, p > .10, \eta_p^2 = .00$; older adults: $F(1.46) = .32, p > .10, \eta_p^2 = .01$) and age heterogeneity (young adults: $F(1.40) = 0.49, p > .10$; older adults: $F(1.46) = 1.36, p > .10$).

To investigate whether the samples of the two studies differed in terms of their intellectual functioning, four separate 2-by-2 factorial univariate ANOVAs⁷ were conducted, one for each test. The dependent variable was the respective test score; age group (young adults vs. older adults) and study (Dynamics vs. CoOP-Mind) were included as between-person factors. The

⁷ A multivariate ANOVA yielded comparable results. The multivariate results can, however, not be so easily interpreted because one Identical Pictures test score for a young adult was missing, meaning that the corresponding sample of young adults in the analysis consisted only of 17 participants.

results are displayed in Table 3. In line with previous results, older adults scored lower than younger adults in the two tests from the domain of cognitive mechanics. Contrasting younger and older adults in the domain of cognitive pragmatics showed a comparable level of performance in both age groups in the Vocabulary test and that the older adults had an advantage over younger adults in the Spot-a-Word test. More interestingly, across all four tests, neither a significant effect of study, nor any age-by-study interaction was found. These findings suggest that the selection effects in the sample used in the analyses for this dissertation are small or non-existent in terms of intellectual functioning.

Table 3. *Examining Sample Selectivity: Contrasting Four Intelligence Tests*

Cognitive Mechanics		
	Digit Symbol Substitution	Identical Pictures
ANOVA Effect		
Age	$F(1,86) = 98.91, p < .05, \eta_p^2 = .54$	$F(1,84) = 188.59, p < .05, \eta_p^2 = .69$
Study	$F(1,86) = 0.45, p > .10, \eta_p^2 = .01$	$F(1,84) = 0.12, p > .10, \eta_p^2 = .00$
Age*Study	$F(1,86) = 0.00, p > .10, \eta_p^2 = .00$	$F(1,84) = 1.68, p > .10, \eta_p^2 = .02$
Cognitive Pragmatics		
	Vocabulary	Spot-a-Word
ANOVA Effect		
Age	$F(1,86) = 0.13, p > .10, \eta_p^2 = .00$	$F(1,86) = 28.52, p < .05, \eta_p^2 = .25$
Study	$F(1,86) = 0.02, p > .10, \eta_p^2 = .00$	$F(1,86) = 0.48, p > .10, \eta_p^2 = .00$
Age*Study	$F(1,86) = 0.35, p > .10, \eta_p^2 = .00$	$F(1,86) = 0.39, p > .10, \eta_p^2 = .01$

4.3 Procedure

The Intra-Person Dynamics study design consisted of three parts: A comprehensive baseline assessment, followed by a period of daily assessment, and a post-test assessment. This dissertation focuses exclusively on variables measured at the baseline assessment and over the daily assessment period. It is only the procedural details of these two parts of the design that are described in the following.

4.3.1 Baseline Assessment

For the baseline assessment, participants underwent a comprehensive assessment battery on three consecutive days covering a wide range of psychological tests and assessments of physical functioning. A trained interviewer distributed questionnaires and written tests within small group sessions (group sizes varied between two and five) and oversaw the participants while they completed them. People trained in the experimental procedure carried out the computerized

assessments of performances in sensorimotor and cognitive tasks in individual sessions. The first two sessions for baseline assessment lasted approximately four hours, while the third session was about two hours long.

At the beginning of the baseline assessment, all participants received an introduction to the study and were shown some examples of the tests and computerized tasks to be used in the study. After this presentation, participants signed a consent form before the actual assessment began. On the first day of assessment, all participants started by completing a questionnaire covering a broad range of personality variables (e.g., the NEO), which lasted about one hour. The questionnaire was followed by a series of written tests assessing spatial abilities, word knowledge (i.e., Vocabulary), and processing speed (i.e., Digit Symbol Substitution). The participants required about 25 minutes altogether to complete them. After these two blocks of group assessments, participants were tested individually on a number of sensorimotor and cognitive functions throughout the first two days. The cognitive tasks consisted of working memory tasks (i.e., verbal and spatial n-back tasks) as well as computerized measures of both fluid “mechanics” (i.e., perceptual speed) and crystallized “pragmatics” (i.e., word knowledge). The sensorimotor tasks included postural control, assessments of visual and auditory sensory acuity, and a variety of clinical tests measuring sensorimotor function (e.g., grip strength, Romberg stand). In addition to the objective performance assessments, the participants filled out a thorough health questionnaire, which included subjective reports of medication, any illnesses they were currently being treated for, and fitness activities amongst other things. The order in which the tasks were carried out differed between participants due to space limitations. Some participants began with the cognitive tasks in one test room and then moved into another room where the sensorimotor tasks were carried out. Other participants completed some of the sensorimotor tests first and then moved to another room where they performed the cognitive tasks. The order of tasks was counterbalanced across participants. The third day of baseline assessment consisted of the further working memory tasks (i.e., digit memory span and word memory span) and the assessment of postural control performance under challenging dual-task conditions (i.e., standing in a semi-tandem position and performing a challenging n-back working memory tasks concurrently). All of the participants then started their daily assessment period in the week following the baseline assessment.

4.3.2 Daily Assessments

In the daily assessment period, a subset of the tests carried out at the baseline assessment was carried out again in individual daily sessions that lasted for one hour. The daily assessment period

ran from Monday to Friday across a nine-week time window, resulting in a maximum of 45 daily assessments for each participant. If a participant missed a day for personal reasons (e.g., unexpected illness), additional assessment appointments or sessions were made for the week following the initial nine-week assessment period. The testing sessions were scheduled between nine o'clock in the morning and eight o'clock in the evening. To account for the possible influence of daily circadian rhythms on intraindividual variability (e.g., West, Murphy, Armilio, Craik, & Stuss, 2002b), the participant was asked to choose a fixed time during the day when the tests should take place. The one-hour daily assessment slots included the same sequence of tasks for every participant on every day. All daily sessions were conducted by people trained in the experimental procedure. In each daily session, the participants first filled out a questionnaire in which they reported their current situation regarding various aspects of their well-being. Their pulse and blood pressure were recorded afterwards. To measure participants' objective performances in sensorimotor and cognitive functioning, they first performed an auditory oddball task for five minutes, which was then followed by two experimental blocks in which they carried out a spatial working memory task. Spatial working memory was assessed using two versions of a spatial n-back task. The easier version was used in the first block of four trials, the second, more demanding version was used in the second block which also contained four trials. Between the two blocks, participants were asked to take a short break of one minute. After assessing cognitive functioning, sensorimotor performance was examined with regard to the individual's level of postural control performance in the semi-tandem position. Postural control performance was tested in two experimental blocks made up of five trials. These two blocks measured how stable the participants were able to stand, both in single-task performance and when carrying out a spatial working memory task while standing (dual-task performance). The order of the single task and dual-task blocks was counterbalanced between persons with regard to age group and gender. To attenuate the influence of fatigue, participants were required to sit down for five minutes between the two measurement blocks. The participants were asked to perform then a Digit Symbol Substitution test during this time. Before each of the sensorimotor and cognitive assessments, the participants were asked to rate how motivated they were to carry out the task and how confident they were of obtaining a good performance. The ratings were made on a seven-point Likert scale. Furthermore, after carrying out each task, the participants were also asked to rate how good their performances had been and how satisfied they were with the outcome, again on a seven point Likert scale. These measures of motivation and subjective performance do not-actually form part of this dissertation, but are analyzed by other members of the Intra-Person Dynamics Project. It was only after completing the rating process that the participants received an objective feedback of their performances. A schema of one particular

design permutation for the daily assessments is provided in Figure 4. In this particular case, single-task postural control was tested before dual-task postural control. The participants earned 10 Euros for each daily session. They received their earnings on a weekly basis every Friday during the daily assessment period.

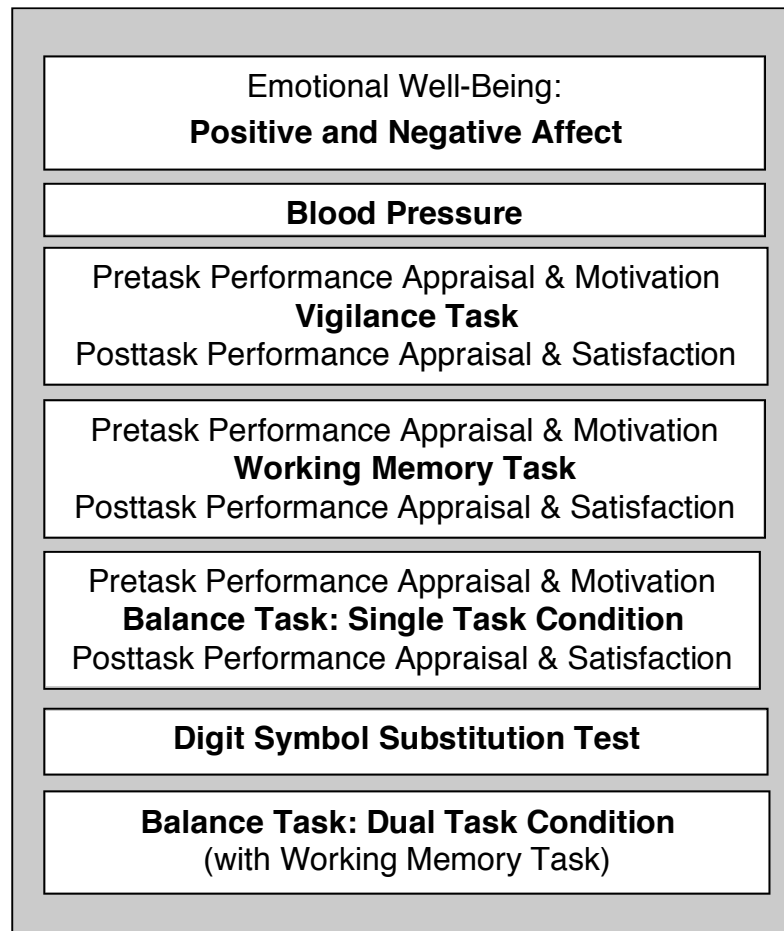


Figure 4. Experimental Design of Daily Assessments.

4.3.2.1 Participation Rate

A number of strategies were implemented to ensure that the attrition rate in the daily assessments was minimal. The consent form signed by the participants, for example, contained a paragraph emphasizing that the participants were willing to attend all 45 daily sessions and can be thought of as an informal contract between participants and experimenters. Furthermore, it was seen as particularly important that the people carrying out the experiment and the participants got on well. In order to help achieve this goal, the number of different researchers assessing a particular participant was kept to a minimum. In addition to this, the principal researchers, predoctoral, and postdoctoral students visited the laboratory regularly and made the participants aware of their vital role in the study. A monetary bonus given only to participants who attended least 40

sessions provided another incentive for regular participation. Finally, other small incentives were provided such as free muffins every Friday, chocolate on Saint Nicholas' Day and Easter, and birthday cards. In addition to the efforts made to motivate participants mentioned above, participants were free to choose, within certain constraints, their preferred testing time, thus allowing them to incorporate the daily assessment sessions into their routine with ease. Table 4 displays the sample frequencies of the daily sessions. The participation rate in the final sample varied between persons but was overall surprisingly high.

Table 4. *Number of Daily Assessments in the Sample*

Daily Sessions (n)	Participants (n)	Participants (%)
42	1	2.8
43	2	5.6
44	3	8.3
45	28	77.8
46	2	5.6

Two participants attended 46 sessions because they were unable to finish the assessment battery on a particular day, which meant these sessions were repeated at the end of the daily assessment period. In the following analysis, their participation rate was set at 100%. The overall participation rate in the final sample was 99.38%. There was no statistical difference in the sample between young adults, which had a participation rate of 99.38% ($SD_{rate} = 1.69\%$), and older adults, which also had a participation rate of 99.38% ($SD_{rate} = 1.56\%$).

4.3.2.2 Time-of-Day Effects

Figure 5 shows the frequency distribution of the daily assessment times.

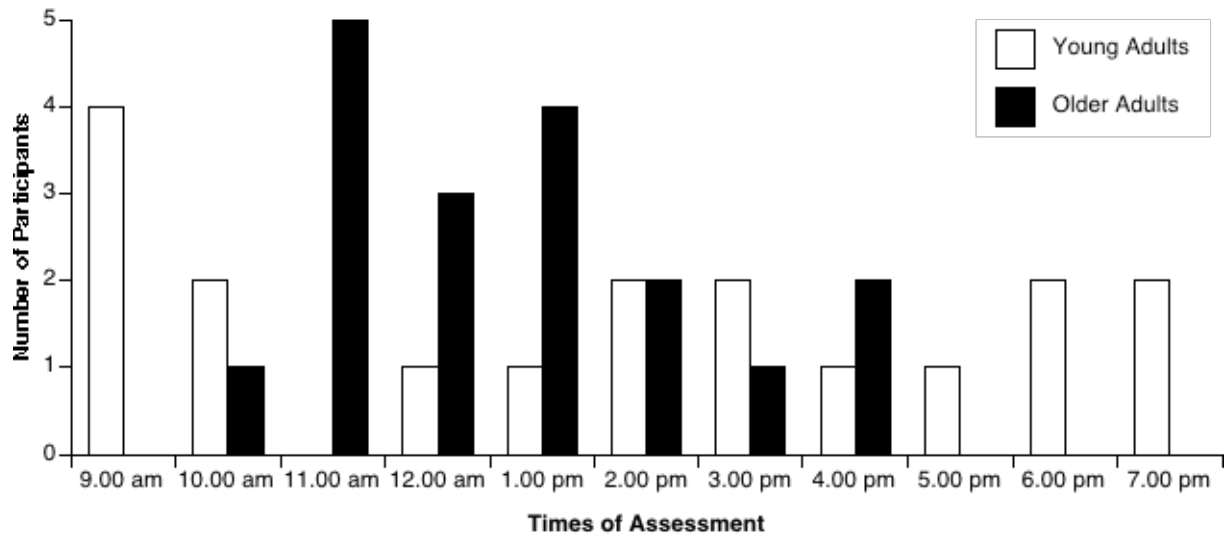


Figure 5. *Times of Daily Assessments among Young Adults and Older Adults.*

The times chosen by young adults did not differ significantly to those chosen by older adults, $F(1,34) = 1.07, p > .10$. The young adults selected testing times relatively evenly from the whole range of available times whereas the older adult group preferred assessment times before and around noon (Levene's test of equality of variances: $F(1,34) = 12.68, p < .05$). The time of the day at which the assessment of psychological functions takes place has been shown to affect results. However, empirically it has been found that older adults in particular perform worse in the evening than in the morning, whereas younger adults remain relatively unaffected by the time of assessment (May, Hasher, & Stoltzfus, 1993; West et al., 2002b). Therefore, it can be concluded that potential age differences are unlikely to be confounded by age differences in daily testing times

4.4 Description of the Tasks and Data Analysis Procedures

This dissertation is concerned with intraindividual variability in postural control performance and within-person interrelations of postural control performance and spatial working memory. Thus, in the following parts of the Method section, a comprehensive overview of all tests and measures obtained in the Intra-Person Dynamics study has not been provided. Only the tests and measurements that are relevant to this dissertation are described in detail.

4.4.1 Postural Control Tasks

Postural control was assessed in two experimental conditions. In both conditions, participants were required to stand on a force platform for 68 seconds. They were supposed to fix their gaze on the middle of a computer screen located about 0.5 m away. The height of the screen was individually adjusted to the eye-level of each participant. In the first condition, termed simple standing from here onwards, participants stood in a semi-tandem standing position and were supposed to fix their gaze on a cross on the computer screen. They were instructed to stand as still as possible. In the second condition, referred to as dual-task standing from here onwards, participants performed a demanding 2-back working memory task while standing on the force platform. The visual stimuli of the task were presented at the same location as the fixation cross in the simple standing condition. In the dual-task standing condition, participants were instructed to stand as still as possible and to perform the working memory task as fast and as accurately as possible. The following sections give detailed descriptions of the measurement devices, the dependent postural control measure, and the data handling procedures.

4.4.1.1 Apparatus

Postural control is effectively measured by assessing the forces that participants apply to the ground to maintain their upright posture (Winter, 1992). The resulting vector of ground reaction forces at a given time is called the center of pressure (COP). The center of pressure was measured with a force platform (60 cm x 40 cm; Kistler force platform 9286AA, Kistler Instrumenten AG, Winterthur, Switzerland), which was connected to a high-speed measurement computer (μ -MUSYCS; m-M-S_Eth-RJ45). 12 sensors built into the force platform measured medio-lateral, anterior-posterior, and vertical components of ground reaction forces and momentums. The measurement computer was used to sample their signals at a rate of 80 MHz and calculated x-y coordinates of center of pressure (COP) positions for every millisecond, while the experimental trials were initiated with a separate experimental computer (NEXOS Pentium III-IV; 500-1000 MHz, PC /NT), which also controlled trial length, measurement onset, and stimulus presentation in dual-task trials.

4.4.1.2 Quantification Parameters of Postural Control

The human upright body can be modeled as a multilink inverted pendulum (Winter, 1995). Postural stabilization requires that the center of mass of this pendulum remains constantly above the base of support. Regulating forces have to be applied to the ground continually to counteract

the force of gravity. Shifts in the center of pressure, therefore, represent the body's neuromuscular responses to movements in the center of gravity in order to ensure postural equilibrium (Winter, 1992; 1995). Postural instability can be quantified by the level of the deviations from a given point of equilibrium. In this dissertation work, this point of equilibrium was estimated by the arithmetic mean of the COP-distributions in directions x and y. In this respect, postural sway increases if deviations from the arithmetic mean of the COP-distributions increase. The area created by the path of the COP throughout the duration of one trial parameterizes postural control performance. Larger areas imply greater difficulties of the postural control system to maintain postural equilibrium. Large areas thus indicate poor postural control and small areas good postural control.

COP areas were computed by using a Matlab program (Matlab 6.5). The postural control time series recorded by the measurement computer was first filtered using a sixth-order low-pass Butterworth filter (cut-off frequency = 6 Hz) to remove high-frequency measurement artifacts. After this smoothing of the data, the arithmetic mean of the distribution was calculated, around which an imaginary circle was drawn. This circle was cut into 360 one-degree slices. The extreme COP-value recorded in each of the 360 directions from the centre of the circle marked the outer border of a given slice. The 360 areas thus obtained were added together to calculate the overall area. Therefore, the area measure used in the analyses combines the deviations from the arithmetic mean in the medio-lateral and the anterior-posterior direction. The area is thus an exact estimate of the extreme deviations from the point of equilibrium. This means it is not an indicator of the average performance but rather of the absolute performance within a given trial. Thus, it provides an accurate and complete estimation of the area spanned by the COP excursions (Krampe et al., 2003). It has been shown empirically that standardized COP-displacement measures such as the standard deviations of displacements are less sensitive to experimental manipulations than to the actual range of motion (Raymakers, Samson, & Verhaar, 2005).

4.4.1.3 Standing Positions

Research has shown that the more difficult a given motor task is, the more cognitive control is required for its effective execution (e.g., Dault, Geurts, Mulder, & Duysens, 2001). Thus, the difficulty of the standing position was manipulated to increase the likelihood of being able to demonstrate interactions between time-varying cognitive efficacy and time-varying postural control performance. In literature on postural control, a tandem-stand position is often used to increase the demand of postural regulation in comparison to the normal standing position

preferred by individuals.⁸ In the tandem-stand position, the feet are placed one in front of the other on a straight line. Pilot studies showed that it is too difficult for some older adults to maintain a tandem stand for a trial length longer than one minute, especially if repeated measurements are involved. Thus, it was decided to adjust the difficulty level of the standing position individually.

For the baseline assessment, participants were instructed to choose a comfortable standing position with their feet apart at a distance approximately shoulder width and with their toes pointing away from the sagittal plane.⁹ The standing position chosen by the participants was marked and used throughout the assessment of the shoulder-width stand at the baseline assessment. After two practice trials, the postural sway in this position indexed by the COP area was measured in four trials and then averaged. The adjustment of the stance difficulty that followed was then based on this average performance. Pilot trials carried out with student assistants had shown that in young adults, changing the stance position from a shoulder-width stand to a tandem stand was associated with an average increase in the COP area by a factor of approximately 2.5 to 3. Therefore, it was attempted to find a stance position for every participant that was associated with an average COP area that was 2.5 to 3 times larger than his or her average performance when in the shoulder-width standing position. To achieve this goal, the stance position of every participant was individually adjusted so that the associated COP area met the criterion described above. Before the beginning of the difficulty adjustment procedure, every participant selected his or her preferred anterior foot. Once selected, the anterior foot was kept constant across the study. The difficulty adjustment procedure began with the participant standing in a full tandem standing position. If the participant's COP area displayed in the tandem

⁸ The average preferred standing position is characterized by a distance of about 17 cm between heel centers, with an angle of approximately 14° between the long axes of the feet (McIlroy & Maki, 1997).

⁹ The empirical literature regarding postural control reports that if participants are free to choose their own preferred standing position, there are large interindividual differences in these positions (Chiari, Rocchi, & Cappello, 2002; McIlroy & Maki, 1997; Mouzat, Dabonneville, & Bertrand 2004). Moreover, McIlroy and Maki (1997) have shown that older adults have a tendency to choose a stance that not only places the feet closer together than the typical stance chosen by younger adults but also has a wider angle between the feet in comparison to younger adults. A narrower distance between the feet is associated with an increased postural sway in younger adults (Chiari et al., 2002; Mouzat et al., 2004). However, it is thus far unknown if the average stance position preferred by older adults indicates an adaptive adjustments to their altered postural control capacities (McIlroy & Maki, 1997). Furthermore, differences in the distance between the feet have to be quite large to affect postural control (Mouzat et al., 2004). The instructions that were given to the participants regarding their shoulder-width stance position tried to ensure a balance between normalizing the feet position for reasons of comparison and the need to simulate a “natural” stance position.

standing position fell within the intended range, the adjustment procedure stopped. If the area in the tandem standing position was larger than three-times the participants area in the shoulder-width standing position, the anterior foot was moved sideways, which increased the base of support and therefore the postural stability. These steps were repeated until the adjusted standing position lead to a COP area within the criterion range of 2.5- to 3-times the area obtained in the shoulder-width standing position.

On average, the individuals' standing positions that were obtained by the difficulty adjustment procedure successfully met the intended criterion. The ratio between the COP areas measured in shoulder-width standing position and in the semi-tandem standing position, which was attuned to every participant, was on average 2.82 ($SD_{ratio} = .53$). The average ratios differed slightly between age groups. The ratios were on average 2.91 ($SD_{ratio} = .57$) among young adults and on average 2.72 ($SD_{ratio} = .49$) among older adults. The small age difference between these ratios was, however, not significant, $F(1,34) = 1.153, p > .10, \eta_p^2 = .04$.

4.4.1.4 Postural Control Assessed at the Baseline Assessment

The postural control performance data collected at the baseline assessment and then used in the analyses consisted of four trials of simple standing and four trials of dual-task standing (i.e., standing and performing the spatial working memory task simultaneously). One older person did not perform dual-task standing in the baseline assessment because of time constraints. The correlation between simple standing and dual-task standing was $r = .56$ ($p < .05$) in the young adult group and $r = .87$ ($p < .05$) in the older adult group. Thus, a COP area for this particular person was estimated with the EM-algorithm (Dempster, Laird, & Rubin, 1977) that is implemented in SPSS.

4.4.1.5 Postural Control Assessed in Daily Sessions

Daily measures of postural control performance included simple standing in the individually adjusted semi-tandem position and a combination of standing in the semi-tandem position and performing the spatial working memory task at the same time. Both postural control performances were assessed in blocks of five trials, each lasting 68 seconds. The order of the two blocks was counterbalanced with respect to age group and gender.

The maximal number of data points per experimental condition over the 45 days was equal to 8050 if the participation rate is taken into consideration (i.e., 36 persons * 45 sessions * five trials per conditions * participation rate). Due to computer problems and other difficulties

that arose at particular daily assessment sessions (e.g., one participant had bruised hands and was unable to push buttons), five trials of simple standing (0.06% of all possible trials) and 25 trials of dual-task standing (0.31% of all possible trials) were not included in the analyses. Furthermore, the data from seven trials (0.09% of all possible trials) of two young individuals were not stored on the hard disc and were therefore lost. For these seven trials, the COP area information was available from written recordings made by the people carrying out the experiment. The postural control performances for these seven missing trials were estimated using the EM algorithm implemented in the SPSS software package (SPSS12.01 for Windows). The prediction was based on all information about daily performances in postural control for both individuals.

4.4.1.6 Identification of Extreme Data in Postural Control

In postural control assessments, participants sometimes forget the task instructions and move an arm or leg or begin to speak especially during repeated measurements. Forgetting the task instructions leads to large COP areas that are not a fair assessment of the momentary postural control capacity of the participant. The people carrying out the tests were instructed to watch out carefully for improper behaviors (e.g., speaking). However, due to the nature of the study (i.e., highly intensive, repeated testing), it cannot be assumed that such behavior was always spotted when it occurred. Thus, potentially extreme data points were further evaluated with respect to the measured COP-areas. An extreme value was defined as a single performance value outside the person's normal range of potential performances.

Dealing with this kind of extreme data is a difficult issue in micro-longitudinal designs. It can be expected that at least some participants show learning over time. Therefore, a particular area value can be a plausible value at the beginning of the learning but has to be considered an extreme value towards the end of the learning curve. Furthermore, one of the main focuses of this dissertation was the investigation of processing fluctuations. There are no strict criteria to decide at which point an extreme value is an indication for low processing robustness or an indication of improper task performance (e.g., the participant moved a leg or spoke during assessment).

The procedure for identifying extreme values was thus based on the time-series of the single trials, as it was intended to identify single trial extreme performance. In the first step of analysis, the single trials of a participant were ordered by time. The first trial on the second day of assessment was, for example, the sixth trial in the time-series, because five trials had already been carried out on the first day of assessment. The particular trend in a given participant's data was a priori unknown. Furthermore, trends in postural control could have differed between participants

and between the two postural control conditions (simple standing and dual-task standing). Therefore, non-parametric locally weighted regressions were fitted separately for every participant and separately for the two conditions (*loess*; Cleveland & Devlin, 1988) to identify extreme values in relation to trends in the data. The loess approach is particularly suitable because of its high flexibility in dealing with interindividual differences in trends. The loess function fitted the individual's average postural control performance at a given point in time (i.e., at particular trial number in the time-series) with a locally weighted moving regression very similar to a moving average in time-series analyses. In contrast to a moving average, the influences of the data points in a loess regression are weighted according to their distance from the point to be estimated. The weighting followed a Gaussian distribution. The closer a specific data point is to the point of interest, the stronger its influence is in the regression. One parameter that can be freely chosen in the loess procedure is the smoothing parameter f , which determines the window size considered in the regression. As the smoothing parameter f decreases, the loess function becomes less smooth and fits the data more closely. It was decided to opt for a narrow window size of .20 (i.e., 20% of the data were considered in the local regression estimates) to achieve a close fit to the empirical data. After the estimation of the individual loess curves, the within-person standard deviations around these curves were calculated. An extreme value was defined as a single-trial performance value that differed by four standard deviations from the person's loess curve. Any extreme value was then excluded from subsequent analyses. As an example, Figure 6 displays the time-series for simple standing of one of the young adults. The average loess trend follows the empirical data relatively closely. The criterion of four standard deviations that separate strong processing fluctuations from outliers corresponds to the average trend.

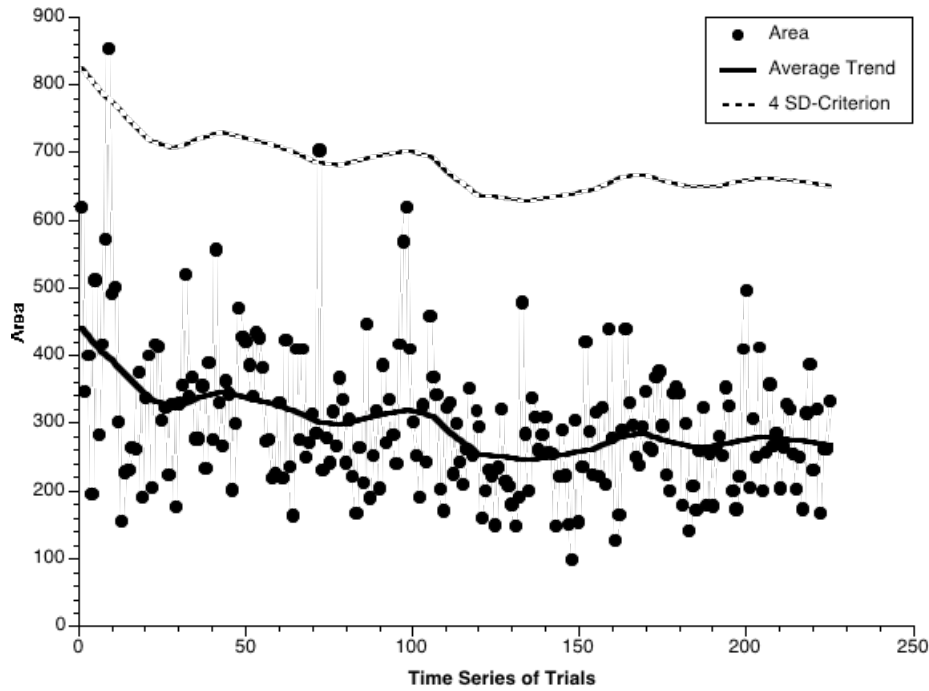


Figure 6. *Illustration of the Extreme-Value Identification Procedure.*

The number of outliers identified was fairly small. In the young adult age group an average 0.27% (SD = .31%) of all data points in simple standing and 0.50% (SD = 0.50%) of all data points in dual-task standing were excluded from the analyses. The number of outliers identified in the older adult age group was 0.45% (SD = 0.40%) in simple standing and 0.42% (SD = 0.50%) in dual-task standing. For any single individual, the number of excluded outliers never exceeded 1.36% in either of the two conditions. A 2 by 2 repeated measurement ANOVA with the percentage of extreme values as the dependent variable and postural task as a within-person factor and age group as a between-person factor revealed no statistical differences between tasks ($F(1,34) = 1.345, p > .10, \eta_p^2 = .04$) and age groups ($F(1,34) = 0.17, p > .10, \eta_p^2 = .01$). The task by age group interaction ($F(1,34) = 2.15, p > .10, \eta_p^2 = .06$) with respect to the amount of extreme values was not significant either. Therefore, given the overall small number of extreme values and the results of the repeated measurement ANOVA, subsequent analyses conducted in this dissertation are unlikely to have been affected by the extreme value identification procedure.

4.4.2 The Spatial Working Memory Task

There are six reasons why a particular working memory task for the investigation of aging-related differences in daily couplings between postural control and cognitive processing was used in this dissertation. First, measures of working memory capacity show a high correlation with psychometric measures of fluid intelligence (Fry & Hale, 1996; Süß, Oberauer, Wittmann,

Wilhelm, & Schulze, 2002). Second, working memory tasks are particularly sensitive to senescent processes (Mayr & Kliegl, 1993; Mayr, Kliegl, & Krampe, 1996; Verhaeghen, Mayr, & Kliegl, 1997). Third, working memory functioning strongly requires executive control (Smith & Jonides, 1999). Fourth, reliable age-related differences in measures of intraindividual fluctuations can be found in the execution of working memory tasks (West et al., 2002). Fifth, it has been shown that the regulation of posture is significantly affected by the concurrent execution of working memory tasks (Dault et al., 2001; Maylor & Wing, 1996; Rapp, Krampe & Baltes, 2006). Sixth, a spatial working memory task was chosen because some studies suggest that spatial cognitive processing has a stronger influence on sensorimotor performance than verbal cognitive processing (e.g., Kerr, Condon, & McDonald, 1985; Maylor, Allison, & Wing, 2001).

4.4.2.1 Task Description and Apparatus

Working memory was measured with a particularly demanding variant of a spatial n-back working memory task (adapted from Kwon, Reiss, & Menon, 2002). The spatial n-back task consisted of encoding, processing, and memorizing a series of spatial locations in sequence, which were presented one after another. Specifically, in each given trial, the participants were presented a dot that appeared in one of the eight outer squares in a 3 by 3 grid (the middle square was not used). Upon seeing the dot, the participants were then supposed to shift the location of the dot on to the neighboring square in the grid, moving in a clockwise direction. They were also supposed to remember this new location. A target was then defined as occurring when the non-shifted location of a new dot was identical to the shifted location of the dot presented two trials before. In each experimental block of trials, participants were shown a random series of 22 stimuli on the computer screen. The random series presented in any given experimental block-by-session combination was identical for every participant. The presentation time for each stimulus was 500 ms, while the inter-stimulus-interval (ISI) was 2500 ms. In both the sitting and standing positions, the participants gave their responses using hand-held button-boxes. By pressing the button held in their right hand, they could indicate that a target had been reached; by pressing the button held in their left hand they could indicate that no target had been reached. Responses faster than 100 ms and slower than 3000 ms were treated as errors. By definition, the first two locations of a series could only be non-targets. While participants were instructed to respond to these stimuli, these responses were excluded from the analyses. In the seated position, the participants' responses were recorded using a National Instruments Data Acquisitions Cards PCI MIO 16E connected to a McIntosh PowerPC 7100, 80 MHz computer. Under the dual-task condition, whereby the working memory task is carried out while standing on the force platform, the

participants' responses were measured using a separate measurement computer (μ -MUSYCS; m-M-S_Eth-RJ45) that was connected to a NEXOS Pentiums III-IV, 500-1000 MHz computer. The sampling rate for the reaction time data was 1000 Hz in both cases.

4.4.2.2 Reaction Time as the Dependent Variable of Working Memory Performance

The analyses of the spatial working memory data is not of primary concern for this thesis. Other ongoing subprojects within the Intra-Person Dynamics project group (S.-C. Li et al., 2005; Schmiedek & Li, 2004) examined the working memory task in more detail. Therefore, the analysis of working memory data is only briefly reported when it is directly relevant to questions regarding the intraindividual coupling between postural control and cognitive performance. The investigation of the coupling between postural control and working memory was restricted to an analysis of each of the two performances assessed in a single-task context. Intraindividual variability in simple standing postural control was related to working memory performances assessed while participants were seated. Analysis of the working memory data demonstrated that the intraindividual variability of measures of accuracy was extremely low in the young adult age group (Schmiedek & Li, 2004). Therefore, the reaction time for correct responses was used as a dependent cognitive variable in the analysis in this dissertation.

4.4.2.3 Assessment of Working Memory at Baseline

Spatial working memory performance was measured at the baseline assessment using three blocks of trials. Each trial lasted 68 seconds. All participants completed all trials.

4.4.2.4 Daily Assessments of Working Memory

On a daily basis, working memory performance was measured using four blocks of trials. The maximal number of experimental blocks obtainable, taking the number of participants and the number of days of assessment, and the participation rate into consideration, was, therefore, equal to 6440 (i.e., 36 persons * 45 measurement occasions * four trials per day * participation rate). Due to computer problems or other difficulties that arose in particular daily assessment sessions (e.g., on one day, a blackout occurred during a daily session), 17 of the 6440 blocks (0.26%) were missing.

One of the main interests of the Intra-Person Dynamics study was to examine processing fluctuations in cognitive performance. In the context of the study, it was decided not to run an extreme value identification procedure as was conducted for the postural control data, but rather

to exclude particular trials based on other response pattern criteria (e.g., response patterns in block of trials that indicated that the participant had shifted the dot in the wrong direction). The rationale of this approach was motivated by the fact that there were no clear-cut criteria suggested in research literature which could separate strong processing fluctuations from extreme values in reaction times. To allow comparisons between repeated measurements, participants had to perform the working memory task in a consistent manner. Therefore, a particular block of trials was excluded from analysis if there was a strong indication that a participant had not followed the task instructions properly. To identify when this had occurred, single responses within trials were reanalyzed. In the specific spatial n-back memory task used, the participants had to rotate every point occurring in the grid one square clockwise in order to complete the task successfully. The reanalysis of the single responses followed two alternative evaluation schemes. The first scheme treated a particular response as correct if the point had not been rotated at all. The second alternative scheme treated a particular response as correct if the point had been rotated anti-clockwise. If, in a particular block of trials, a participant gained 25% accuracy according to one of the alternative evaluation schemes in comparison to the usual evaluation schemes (indicating that at least a quarter of the stimuli were not processed in the instructed manner), this block of trials was excluded from the analyses. Following this procedure meant that 1.51% (SD = 2.65%) of all trials of the young adult group and .22% (SD = .43%) of all trials in the older adult group had to be excluded from analysis. The difference between the age groups in terms of the percentage of trials excluded was significant, $F(1,34) = 4.19, p < .05, \eta_p^2 = .11$. The overall percentage is, however, relatively low. Therefore, subsequent analyses are unlikely to be biased by this exclusion procedure.

4.5 Methodological Issues and Analytical Approaches

In the past, research on cognitive and sensory/sensorimotor aging was mostly concerned with the investigation of performance differences in mean level and correlational analysis of between-person differences in level and intraindividual change. In contrast, there have been relatively few studies concerning age differences in intraindividual variability, in particular regarding intraindividual variability occurring on a day-to-day time scale. Consequently, there is as yet no standard way of analyzing intraindividual-variability data. The following sections provide some information about the issues associated with the study of interindividual differences in intraindividual variability. Specifically, the focus here is to link these general issues to the specific problems of the data set at hand in order to motivate the rationale of the data analytic approaches used in this dissertation.

4.5.1 Description of Multilevel Models

The data-analytic approach taken in this thesis is person centered (Bergmann, Magnusson, & El-Khoury, 2003; Molenaar, 2004). The structure of the empirical data used in this regard is made up of 18 young and 18 older individuals that were assessed on approximately 45 days. Hence, the data were obtained from a multi-level sampling scheme, in which participants were sampled first, and days were sampled within participants. As a consequence, the observations are not all independent because daily assessments are nested within participants. Using standard statistical methods such as repeated measure ANOVAs or any other method based on ordinary least squares is problematic with nested data. These methods assume that the observations are independent and underestimate respective standard errors by ignoring the correlated data structure of nested designs. In contrast, multilevel random coefficient modeling or multilevel modeling (MLM) is a highly accurate methodological tool to examine the type of data structure under consideration (e.g., Goldstein, 1995; Nezleck, 2001; Raudenbush & Bryk, 2002; Sliwinski & Bushke, 2004). Specifically, MLM has an advantage over other techniques based on ordinary least squares (OLS) because it is capable of estimating random errors simultaneously and independently at multiple levels of analysis using maximum likelihood estimation.

4.5.1.1 General Principles of Multi-Level Modeling

In multilevel models, a regression equation is estimated for each level of analysis. The coefficients at one level of analysis become the dependent variables in the next level of analysis. Regarding this dissertation, the regression coefficients describing the first level of analysis (i.e., the intraindividual daily time-series) were thus further analyzed at the second level, which was constructed by the differences between persons in these coefficients. In two-level MLM analyses, up to two types of parameters are estimated for every coefficient. The fixed parameters index the average trend or the central tendency of a given coefficient. In this dissertation, a fixed parameter is an estimation of the average coefficient across all intraindividual time-series and is tested against the H_0 of being zero in the population. The random parameters or random error terms either denote the variance unaccounted for by the first level regression equation at the first level (i.e., the within-person level) or represent the interindividual variances and covariances in regression coefficients at the second level (i.e., the between-person level). The latter variance can then be statistically tested in order to ascertain the potential significance of interindividual differences in the coefficients describing the intraindividual time-series. Cross-level interactions indicate that the strength of a coefficient at level one is influenced by between-person differences at level two. An important benefit of multilevel models is that they are robust against unbalanced

designs or missing data. In the full model, the influence of a particular first-level coefficient on second-level parameters is weighted with regard to the number of first-level units that it was computed from (Raudenbush & Bryk, 2002). The small between-person differences in the number of assessments that are part of the empirical data of this dissertation were, therefore, explicitly accounted for in the multilevel analysis.

One assumption of multilevel random coefficient models is that random errors at the first level are normally distributed with a mean of zero and an unknown common variance across all level-two units. Random errors at level two have to be independent from level-one errors and have a multivariate normal distribution with a mean of zero. Simulation analyses have demonstrated that the parameter estimation in multi-level models is robust against violations of the assumption that the error variances are distributed according to the normal distribution (Maas & Hox, 2004a; 2004b). However, the estimated standard errors of the parameters are biased if the sample size is small. The other assumptions here, which are identical to the common assumptions of multiple regression analysis, are that the predictors are fixed and the relationships linear (Hox, 1998; 2000).

4.5.1.2 Testing Hypotheses with Multi-Level Models

The standard statistical test, called the Wald test (Wald, 1943) is used for testing the significance of coefficients in multilevel models. The Wald test divides the estimated parameter by its standard error and compares the resulting ratio with the standard normal distribution. This means, for example, that an estimated parameter statistically differs from zero at the 5% level if the ratio of the parameter estimate divided by its respective standard error exceeds 1.96. The estimated standard errors are only asymptotic (i.e., valid) when sample sizes are large (Hox, 1998). Another necessary condition for the application of the Wald test is the rather unrealistic assumption that the first-level coefficients are normally distributed (e.g., Hox, 1998; 2000). It is clear that with increasing sample sizes at all levels, estimates and their standard errors become more accurate. For practical reasons, the number of participants was limited to 36 people (i.e., level-two units) and the number of assessments was restricted to 45 days (i.e., level-one units). Simulation studies suggest that this data structure allows relatively good estimations of the fixed effects (i.e., mean within-person effects) but also implies a poor level of accuracy in the estimation of the random effects (i.e., variances and covariances of the between-person differences) (Hox, 1998; Snijders & Bosker, 1999). To counteract some of the limitations of the data set at hand, the statistical tests of parameters were carried out using nested model comparisons. In nested model comparisons, the parameter comparisons are introduced

successively. The overall fit of the models is estimated with a full maximum likelihood method. The difference in the deviance statistics (computed as: deviance = -2 the log likelihood) between two nested models is tested with a chi-square test with one degree of freedom. This type of statistical test has been considered to yield more accurate results than the Wald test given limited sample sizes, especially for random model parameters (Goldstein, 1995).

As pointed out by Snijders and Bosker (1999), the statistical test of a cross-level interaction (i.e., the strength of a coefficient at level one is influenced by a level-two variable) has, in most cases, more power than the test of a random variance. A cross-level interaction considers the first level and second level data simultaneously, whereas a test of a random variance at level two is based purely on the between-person differences. Given the relatively small sample size for the level-two analysis and the relatively high level of level-one data density in the data set, a-priori theory-guided cross-level interactions were tested even in the absence of significant second-level random variances.

4.5.2 Dissociation of Processing Fluctuations from Trends

In the theoretical background to this dissertation, processing fluctuations were defined as short-termed maladaptive deviations from a potential performance. To investigate day-to-day performance fluctuations, participants were assessed repeatedly over many days. It was unrealistic to assume that the potential level of level of performance would be constant across the whole time frame. Instead, the performance was likely to be influenced by the amount of learning experience or possibly by drops in motivation. A specific approach was employed to dissociate processing fluctuations from long-term, trends in performances. The potential shapes of the within-individual time-series of the postural control tasks and the working memory task were unknown at the outset of the study. As a consequence, the procedure of dissociating processing fluctuations and trends was relatively data driven in nature. However, preliminary analyses were carried out first to describe the effect of learning on within-person changes in performance with theoretical functions (e.g., exponential functions). In this case, it was possible to interpret the parameters of the theoretical learning functions in a meaningful way. When this theoretically driven method could not be applied, multi-level models were used to fit polynomials to the within-time series. The exploratory multi-level approach allowed for the simultaneous estimation of the average trend and interindividual differences in trends.

4.5.2.1 Theoretical Practice Functions

The effect of practice on performances is commonly expressed as the negatively accelerated influence of the number of learning experiences on the level of performance (but see Ram et al., 2005). Characteristically, strong learning gains can be observed at the beginning of a series of repeated assessment occasions. The strength of these learning gains flattens out in the course of consecutive learning experiences until a process specific limit is reached asymptotically. This has been found consistently across a large variety of sensorimotor, perceptual, and cognitive tasks (see Heathcote, Brown, & Mewhort, 2000; Ritter & Schooler, 2001, for review).

Selecting the appropriate function to model the shape of learning curves (i.e., the function relating learning gains to the number of learning experiences) has been a contentious issue in the study of learning. In the literature, exponential and power functions in particular have been contrasted in terms of their explanatory value. In their most common form, these two functions have three parameters. The range of the learning parameter describes the overall improvement of performance. The asymptote indexes an upper boundary of performance. The learning rate indicates how fast the asymptote is reached. Both functions differ with respect to the role of the learning rate. Exponential learning functions assume that the rate of learning remains constant over time. A power function implies that the learning rate decreases over time.

It has been shown empirically that power functions provide a very good fit for learning data based on aggregates at the group level. This particular advantage of power functions is likely to be a by-product of averaging non-linear individual learning functions that differ in shape between persons (Myung, Kim, & Pitt, 2000). This has been demonstrated both in simulation studies and by contrasting power functions and exponential functions in terms of how well they can fit single individual learning curves. Moreover, in extreme cases, the shape of the averaged learning curves does not resemble the shape of a single one of its constituent within-person learning curves (Myung, et al., 2000; Haider & Frensch, 2002). In a reanalysis of 40 data sets, including a broad variety of cognitive tasks, exponential curves provided a superior fit for within-person learning curves in comparison to power curves, although power curves were able to provide a better approximation for aggregated curves (Heathcote et al., 2000). Essentially, if one is interested in the influence of learning on performance at the level of the individual, the only viable alternative is to analyze the influence at that level (Heathcote et al., 2000; Schmiedek & Li, 2006).

4.5.2.2 Examples of Trends across Days

In this dissertation, general trends were analyzed with respect to the within-person, individual time-series. The performances in postural control and working memory were aggregated across

trials within each day. The daily means were ordered with respect to when they were recorded. The first step was to plot all the individual time-series and to evaluate their individual shapes by visual inspection. Figure 7 shows four examples of individual time-series over the days of assessment for each of the three daily variables of interest. One young man, one young woman, one older man, and one older woman were randomly drawn from the sample. Figure 7a displays their time-series in simple standing and Figure 7b shows their trends in dual-task standing. Figure 7c shows their within-person time-series of working memory reaction time, which were measured while the participants were seated. These graphs are intended to provide an overview of the interindividual differences in the general trends. It is important to note that the scaling of the y-axis differs between plots.

The examples of time-series shown in Figure 7 illustrate an important finding. The shapes of the within-person trends differ between the domains of functioning. In the sample as a whole, it was found that all participants showed clear learning patterns in working memory. In contrast, the trends in postural control performance showed a greater level of diversity across participants.

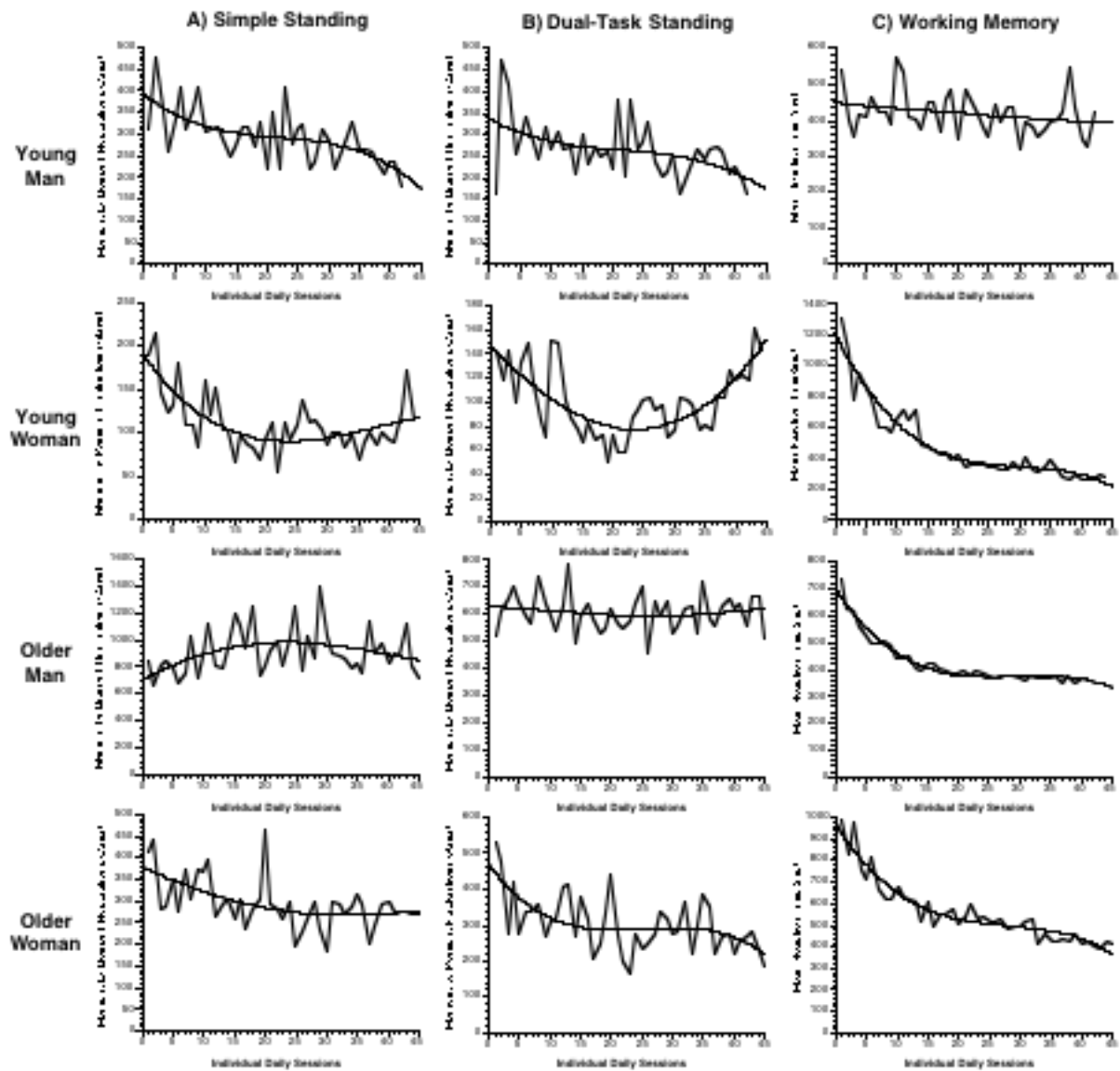


Figure 7. Trends across Days of Four Individuals in A) Simple Standing, B) Dual-Task Standing, and C) Working Memory Speed.

4.5.2.3 Controlling for Intraindividual Trends in Spatial Working Memory

The majority of the within-individual trends in spatial working memory showed the typical pattern found in cognitive learning literature. They demonstrated large learning gains at the beginning of the assessment followed by decreased gains over the assessment period. Therefore, it was decided to fit exponential curves to every individual's time series of the working memory reaction time data using the NLIN procedure in SAS (SAS 9.1 for Windows). The validity of intraindividual curve fits was evaluated using a number of criteria. Statistical criteria such as the contrast between the explained variance and the unexplained variance (i.e., conventional F-test) or significance tests of the three parameters (i.e., range of learning, asymptote, and rate) are not very much informative in this respect. Standard curve fitting procedures do not take the

dependency of observations into account and consequently underestimate standard errors. Therefore, the first step in evaluating the curve fits was to note whether the iteration algorithm converged and, in this way, to judge the plausibility of the obtained parameter estimates, as the parameter estimation is unbiased by data dependency (Hox, 1998). More importantly, the time series of predicted values and the time series of the observed data were plotted against each other and the goodness of fit was evaluated visually. The evaluation revealed reasonable curve fits for the majority of participants but not for all of them. Reanalyzing the data demonstrated that large learning gains were made within the first assessment sessions. Therefore, the exponential curve fitting procedure was carried out on intraindividual time series based on single trials (e.g., the second trial on the second day was the sixth data point in the time-series). This approach resulted in reasonable parameter estimations for all but two participants. The learning trends of one young adult and one older adult were better approximated with a linear trend and not with an exponential curve. Other research has shown that some particular individuals display linear learning trends in cognitive tasks (e.g., Ram et al., 2005). The exponential curve fitting procedure is, however, so flexible that it can approximate a linear trend by choosing high values for the asymptote and the range of learning parameters and setting the learning rate to zero. Contrasting the time series of predicted values and observed data showed good fits for all participants.

4.5.2.4 Controlling for Intraindividual Trends in Postural Control Performance

Inspecting the graphs of intraindividual trends over different days revealed large interindividual differences in their shapes. The dominant pattern was that learning, defined as gains in performance across time, took place. There were, however, also a significant number of participants who displayed a constant or declining performance over the period of assessment. As a consequence, applying a theoretically appropriate function, such as the exponential curve, in order to dissociate processing fluctuations from general trends could not be justified. To obtain sufficient flexibility in order to account for the variety of the observed trends, polynomial curves were fitted to the intraindividual time series. When using this approach, the dissociation of processing fluctuations and general trends is dependent on the order of the polynomials fitted to the data. Multi-level modeling was used both to estimate the order of the polynomial that would describe intraindividual trends sufficiently and to take interindividual differences into consideration. One advantage of polynomials in this context is that the aggregation of polynomial curves of a specific order results in an average curve that is a polynomial of the same order (Estes, 1956; cf. Schmitz, 2000). For the multi-level case, this general law states that the order of the polynomial indexing the average trend (i.e., the second level fixed effect) is in close correspondence with the shape of the individual's trends (i.e., the first level units). Furthermore,

interindividual differences in these trends (i.e., second level random effects) can be modeled within the same analysis.

4.5.3 Contrasting Within-Person and Between-Person Correlations

The focus of the investigation of the hypothetical aging-related increase in the strength of the interrelation between processing fluctuations in sensorimotor function and cognition is on within-person processes. However, as detailed in the Theory section, it would have been premature to assume variance equivalence in the study of age differences in the coupling of both domains. As a consequence, between-person analysis is not a priori informative about the strength or the direction of the intraindividual relationships. The hypothetical data displayed in Figure 8 illustrates this point. Figure 8a displays a between-person correlation between two variables X and Y. Five persons are measured for both variables on a single day. The between-person correlation between X and Y is strong and positive. Figure 8b shows an extreme case of the possible patterns of intraindividual correlations of the same five individuals. The single data point for each individual, upon which the between-person correlation is based, is now part of a sequence of six repeated assessments of every individual for both variables. Assessing these individuals repeatedly on six occasions revealed that all within-person correlations are negative and differ in strength. Furthermore, Figure 8b also shows that as the value of the variable x increases, the relationship between X and Y becomes stronger. This demonstrates that although within-person relationships may be heterogeneous, the existing heterogeneity in the sample at hand may follow a pattern.

Looking at Figure 8 reveals that in the absence of variance equivalence, the between-person correlations are unrelated to the average strength, the direction, or the interindividual differences of intraindividual correlations. Therefore, the first priority of this dissertation was to investigate within-person interrelations between day-to-day performance fluctuations in postural control and spatial working memory. Only then can between-person comparisons be conducted with respect to these interrelations. More specifically, this dissertation seeks to investigate whether between-person variance in within-person relationships is associated with age. Multilevel hierarchical models are well suited for this because they specifically allow both levels (i.e., the within-person level and the between-person level) to be investigated within one analysis.

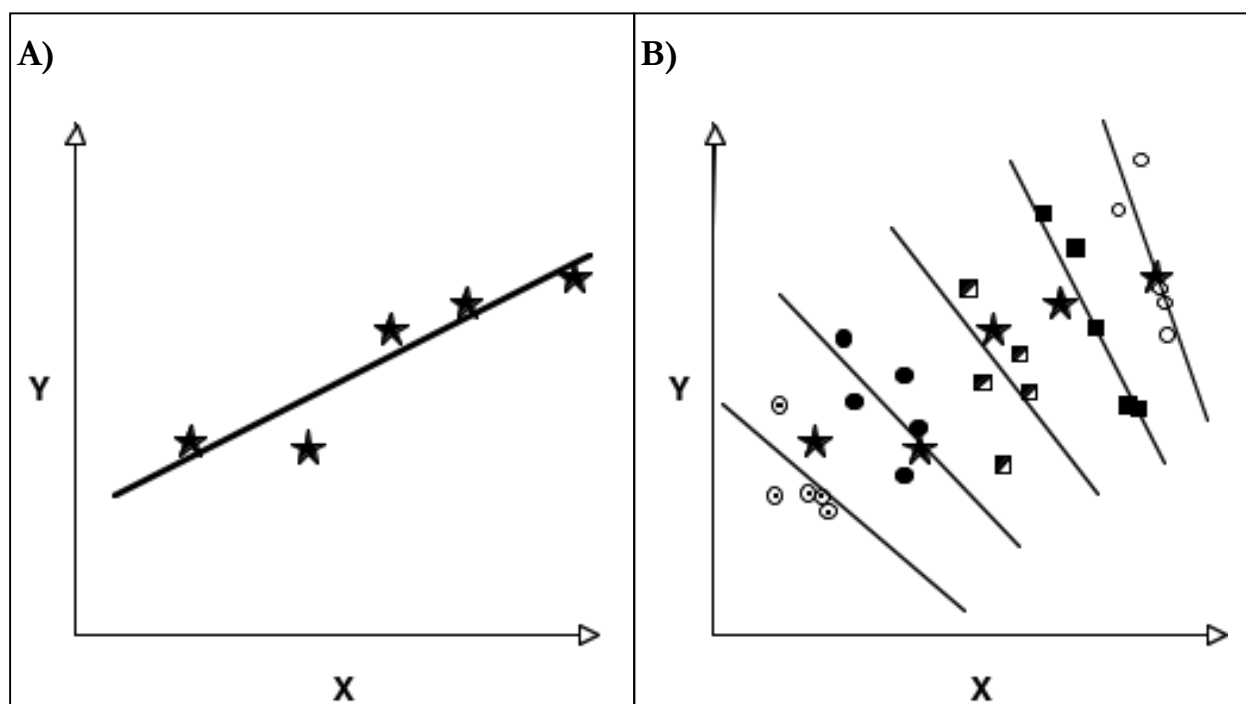


Figure 8. *Illustration of the Independence of Within-Person and Between-Person Relationships: A) Between-Person Relationship based on Five Persons. B) Five Within-Person Relationships.*

4.5.4 Controlling for the Mean in Age Contrasts of Intraindividual Variability

Among researchers in the field of aging, it is generally agreed that the investigation of aging-related differences in intraindividual fluctuations is interesting in its own right. Empirically, however, measures of intraindividual variability and central tendency are often highly correlated. In cognitive aging research, there is concern that aging-related differences in intraindividual processing fluctuations may be caused by a mere statistical phenomenon associated with an aging-related decrease in the average performance (e.g., Salthouse & Berish, 2005). A number of the measures conventionally used in research on aging in the sensorimotor and cognitive domain have a lower boundary for optimal performance. If a variable has a lower boundary, then an upward shift in the distribution of scores, which is indicated by a shift in the mean of the distribution, is accompanied by an increased distance to that boundary. This greater distance, in turn, provides more room for variability. The increase of the within-person processing fluctuations associated with aging could, therefore, simply be a side effect caused by the increasing distance of the average performance from a lower boundary.

There are two common methods that have been used to account for the influence of age differences in the central tendency on age-differences in processing fluctuations. To account for aging-related differences in average performances, the first method partials out age from the repeated measurements before the intraindividual variability is computed (e.g., Hultsch & MacDonald, 2004). This approach is problematic because it accounts for a variable that is

correlated with the mean but does not directly address the problem regarding the distance to the boundary (Salthouse & Berish, 2004). Within the second approach, interindividual differences in mean performances are partialled out directly. The issues regarding this approach are subtle and are discussed in the Discussion section of this dissertation. However, in the Theory section, it was argued that aging-related differences in average performances and processing fluctuations might partly be caused by the same mechanisms. For example, neuronal network simulations demonstrated that a single process parameter, which simulated the effect of deficient neuromodulation, accounted not only for age differences in mean performance but also for variability in performance (S.-C. Li & Lindenberger, 1999; S.-C. Li et al., 2000). In a similar vein, the structural losses in the postural control system that are associated with aging lead on the one hand to higher postural sway (i.e., moment-to-moment fluctuations). On the other hand, they also lead to a higher vulnerability to any perturbations that may occur on different time-scales and therefore to higher trial-to-trial and day-to-day performance fluctuations. Thus, from a theoretical point of view, processing fluctuations and the different time-scales on which they occur are directly related. Moreover, the elementary indicator of the average postural control performance is already a measure of intraindividual variability. Therefore, the question remains as to whether intraindividual variability on longer time-scales (i.e., trial-to-trial and day-to-day fluctuations) contains information that is not captured by intraindividual variability on the moment-to-moment time-scale.

To address this issue, a series of repeated measurement ANOVAs was conducted. First, the age differences in moment-to-moment, trial-by-trial, and day-to-day processing fluctuations were tested separately. In a second step, another series of repeated measurement ANOVAs was conducted in which processing fluctuations on long time-scales formed the dependent variable and moment-to-moment processing fluctuations were integrated as a covariate. The effect of age was examined at the same time.

4.6 Overview of Analyses

This section outlines the analysis conducted in the course of this dissertation. First, the dependent variables are described, followed by an introduction to the actual analysis procedure. This introduction approximately follows the order of the hypotheses as outlined in Section 3. To begin with, the analysis procedure regarding processing fluctuations in postural control on three different time-scales is outlined. Then the analysis procedure responsible for examining the coupling between daily fluctuations in postural control and working memory is also described.

4.6.1 Dependent Variables

The dependent variables used in the different stages of analysis either refer to performance measures that were assessed at the baseline assessment or over the daily assessment period. The dependent variables indexed in postural control performance are introduced first, followed by a description of the dependent variables that indicate spatial working memory performance.

4.6.1.1 Baseline Postural Control Performance

Two dependent measures were used in the analysis process that indexed the postural control performance at the baseline assessment. Both measures indexed moment-to-moment processing fluctuations in postural control at the baseline assessment. The averaged area across the four trials for each condition has been termed the “*baseline level of moment-to-moment fluctuations*” in the following, for the simple standing or dual-task standing conditions respectively.

4.6.1.2 Multiple Time-Scales of Intraindividual Performance Measures of Postural Control

Postural control was assessed for every participant using five trials for each of the two experimental conditions (i.e., simple standing and dual-task standing) every day on approximately 45 occasions.

The shortest time-scale is governed by processing fluctuations that occur from one moment to the next. The average individual level of moment-to-moment processing is indexed by the mean COP area across all daily postural control assessments and has been termed the *daily average level of moment-to-moment fluctuations*. Trial-to-trial processing fluctuations constitute the next longer time-scale. The standard deviations across the five trials for each day were first calculated separately for both conditions. The second step consisted of averaging these standard deviations over all days of assessment. This average has been termed the *level of trial-to-trial fluctuations* in the following. In order to parameterize day-to-day performance fluctuations, COP areas across the five trials were averaged within days. These daily means have been termed the *within-day moment-to-moment fluctuations* in the following. The time-series of within-day moment-to-moment processing fluctuations is not only an expression of day-to-day processing fluctuations but also of trends such as learning effects. Polynomials were fitted to every individual time-series separately for each condition. Deviations from the polynomial trends indicated processing fluctuations. A negative deviation from the curve, for example, indicated a “good” day in terms of postural control performance because the person performed better on this particular day than might be expected from his or her general trend. The intraindividual deviances from the individual specific curve are

referred to as the *day-to-day processing fluctuations*. The averaged absolute deviations from the participant's trend indicated the level of pure processing fluctuations and have been termed the *level of day-to-day processing fluctuations*.

4.6.1.3 Intraindividual Daily Working Memory Performance Measure

The investigation of processing fluctuations in spatial working memory performances are not the primary concern here but form part of the ongoing work of other members of the Intra-Person Dynamics project group. This dissertation, however, seeks to establish whether day-to-day processing fluctuations in spatial working memory performance are coupled with day-to-day processing fluctuations in postural control. The within-person time series of repeated measurements form the basis for this investigation. In order to obtain a pure measure of processing fluctuations in spatial working memory that was independent of any performance changes due to learning, exponential curves were fitted to every intraindividual time-series. By examining the data, it became clear that learning took place both on and between individual days of assessment. Therefore, the exponential curve fitting procedure was done on intraindividual time-series based on single trials (e.g., the second trial on the second day was the sixth data point in the time-series). The deviances from the individual specific curves were aggregated within days. This series of means that were obtained thus expressed the *day-to-day processing fluctuations*.

4.6.2 Analyses of Group Differences in Processing Fluctuations in Postural Control

The hypotheses under discussion in this dissertation make the assertion that age and sex differences exist on the level of processing fluctuations in postural control. The analysis of group differences is outlined first with respect to moment-to-moment fluctuations, then with regard to trial-to-trial fluctuations and finally with respect to fluctuations between days.

Moment-to-moment processing fluctuations were analyzed with respect to age and sex differences at the baseline assessment. They were also analyzed with respect to their average value over the 45 days of daily assessment. The moment-to-moment processing fluctuations recorded at the baseline assessment correspond to the postural control performance indicators commonly used in the field. However, the average level of moment-to-moment processing fluctuations across the daily assessments is more reliable because the average of 45 different performances over the 45 different days of assessment has been taken. Furthermore, it is to be expected that this value may have been influenced by learning processes.

The trial-to-trial processing fluctuations were at first analyzed without including a control variable but the analysis of age differences in these fluctuations was later controlled for

interindividual differences in moment-to-moment fluctuations at the baseline assessment and across the daily assessment period. Furthermore, within-person relationships between trial-to-trial fluctuations and moment-to-moment fluctuations were estimated using a multi-level model. The final step involved using a regression of trial-to-trial processing fluctuations on moment-to-moment fluctuations across days within individuals. The residuals of these regressions were examined with respect to possible age and sex differences. To estimate the level of day-to-day processing fluctuations in postural control, it was necessary to separate these from trends. In this respect, a multi-level model of time was employed that was based on polynomials in order to explore the optimal amount of control for trends in the daily postural control data. The average absolute deviance from the intraindividual polynomial trends across days indicated the level of day-to-day processing fluctuations and was examined with respect to age and sex differences. In subsequent analysis, the age differences were controlled for interindividual differences in moment-to-moment fluctuations assessed at the baseline assessment and the average performance over the daily assessment period.

Contrasting postural control performances in simple standing and dual-task standing tasks has not formed the focus of this dissertation. Therefore, the analysis was conducted separately for both experimental conditions. The aim of this was to demonstrate that potential findings hold under different experimental postural control conditions.

4.6.3 Analyses of Intraindividual Couplings across Domains of Functioning

In the Hypothesis section, it was predicted that older adults and males would demonstrate a stronger coupling between postural control and attentional control than young adults and females. Fluctuations in postural control were indicated by fluctuations in a spatial working memory task that placed high demands on attentional control. In particular, it was investigated whether day-to-day fluctuations in spatial working memory assessed in a seated position were able to predict day-to-day fluctuations in simple standing postural control. The analyses were restricted to simple standing postural control because it is the least confounded form of analysis that can be used to show cognitive involvement in postural control. In dual-task standing, a potential cross-domain coupling could be a mere consequence of the experimental design. Furthermore, in dual-task standing two outcome measures have to be considered. If participants perform the dual-task under resource competition conditions, performance trade-offs could be expressed either in the postural control performance, in the cognitive performance, or both. The domain in which these trade-offs occur could potentially vary between days.

Couplings between day-to-day changes in simple standing and working memory performances were analyzed using multi-level models. This involved examining within individuals whether day-to-day processing fluctuations in working memory were significantly related to day-to-day fluctuations in postural control after having been controlled for trends. Subsequently, interindividual differences in these relationships were evaluated in relation to age, sex, and the overall level of moment-to-moment fluctuations in postural control and mean working memory performance across days.