5. CENTRAL STUDY

This chapter deals with the main part of this dissertation project – the experimental investigation of potential age-related differences in the positive-negative disparity of emotional memory. As outlined in Chapter 3, this dissertation was designed to examine recent suggestions about a positivity effect in older adults' memory (e.g., Charles et al., 2003) and to disentangle explanations for such an effect. To do so, I examined the impact of differential processing prioritization for emotionally-toned words in two conditions that differentially afford processing prioritization. Specifically, in a multi-trial free recall task, I investigated the recall performance of young and older adults for the same sets of words presented either in a context with strong cues for differential processing (i.e., emotion-heterogeneous list context) or a context with no cues for differential processing (i.e., emotion-homogeneous list context). In addition, the design of the experiment allowed me to investigate the influence of learning, long-term retention, person characteristics, and word characteristics on remembering emotionally-toned words.

The chapter is divided in two main parts: a method section and a result section. The method section offers a detailed description of the design, participants, materials, and the procedure of the experiment. The result section is organized to answer the main research questions of this dissertation project followed by more specific analyses to control for potential confounding variables. The results are summarized briefly and discussed in detail in the general discussion (see *Chapter 6: General Discussion*).

5.1 METHOD

5.1.1 Design

One goal of this dissertation project was to make a contribution to sorting out a memory capacity explanation and a selective attention explanation for potential age-related differences in emotional memory. That is the question of whether the proposed positivity effect in older adults' memory is due to differences in the memory capacity for positive, negative, and neutral materials, or whether older adults show an emotion-based processing priority for positively-toned material. To do this, I used a heterogeneity-homogeneity list paradigm. In this paradigm, different list compositions were constructed that either constrained processing prioritization in homogeneous lists (emotion-homogeneous compositions of either positive, negative, or neutral words) or that facilitated the use of emotion-based processing prioritization in heterogeneous lists (emotion-heterogeneous compositions of positive, negative, and neutral words).

To ensure that both conditions comprised the same sets of words, three list compositions were generated for each condition: three heterogeneous lists consisting of 10 negative, 10 neutral, and 10 positive words, and three homogeneous lists of either 30 positive words, 30 negative words, or 30 neutral words. Equal numbers of young (20-30 years) and older adults (65-75 years) were recruited across and within these conditions. Consequently, 2 (age groups) x 6 (list compositions) = 12 different between-subjects groups were required in the experiment. Within each group, 12 participants were recruited resulting in a total sample size of 144 participants. Additionally, to control for potential sex-related differences and education-based differences in the processing of emotional information, the sample was stratified by sex and educational level. Table 16 presents the sample design in the experiment. Each cell included three persons without a high school degree and three persons with a high school degree.

Table 14

Sample Design of the Experiment

	Young Adults		Old Adults		
Lists	Women	Men	Women	Men	Total
Homogeneous List Condition					
Negative List	6	6	6	6	*24*
Neutral List	6	6	6	6	*24*
Positive List	6	6	6	6	*24*
Heterogeneous List Condition					
Mixed List 1	6	6	6	6	*24*
Mixed List 2	6	6	6	6	*24*
Mixed List 3	6	6	6	6	*24*
	36	*36*	*36*	*36*	*144*

To ensure sufficient acquisition of the words and to examine learning effects, five learning trials were given for each list. After each presentation of the word list, participants were asked to recall as many words as possible resulting in five free recall trials during the learning phase. To my knowledge, this is the first study investigating learning effects for emotionally-toned material. In addition, to examine long-term memory for emotional material, two different retention intervals were used in a following retention phase: (a) a 1-

hour retention interval and (b) a 1-week retention interval. Within each retention interval, participants were asked three times in a row to recall as many words as possible. This procedure of three recall trials following immediately after the other was used to guarantee maximal retrieval of the to-be-remembered words. All trials in the learning phase and in the retention phase were manipulated on the within-subject level. Figure 5 presents the within-subject design of the learning and recall trials.

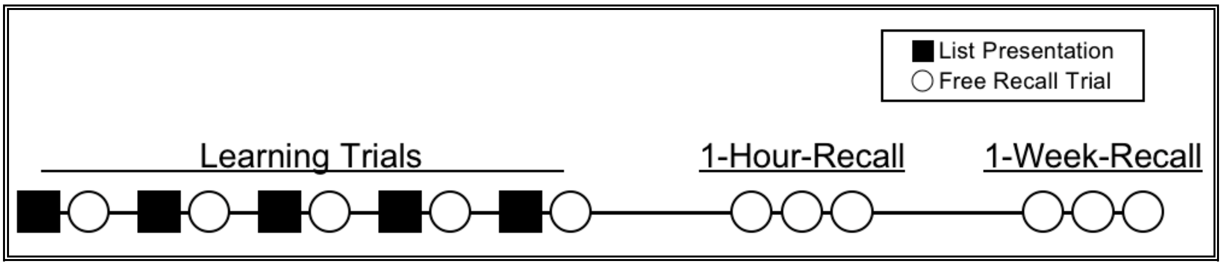

Figure 2. Design for the List Presentation and the Free Recall Trials in the Learning Phase and in the Retention Phase

5.1.2 Material

5.1.2.1 Computer and Software Program

The experiment was conducted on five Macintosh iBook computers (resolution of 800 x 600, 14”). The iBooks were equipped with a 500 MHz PowerPC G3 processor and the operation system Mac OS 9.2.2.

For the experiment, a software program was written in C++. This program controlled the presentation of the words and the selection of appropriate word lists, time intervals, and instructions for each participant by means of an identification number. During the learning task, all instructions were given on the computer screen. To-be-remembered words were presented in black 64-point Geneva font on a uniform white field in the middle of the screen.

5.1.2.2 List Compositions

To address the question of whether the proposed positivity effect in older adults’ memory is due to general differences in memory for positive, negative, and neutral materials, or whether older adults show an emotion-based processing priority for positively-toned material, list composition was systematically manipulated within the heterogeneity-homogeneity list paradigm. For this purpose, I constructed three emotion-homogeneous lists of only negative, positive, or neutral words and three emotion-heterogeneous lists of negative, positive, and neutral words.

Based on the word rating study, a final item pool of 30 negative, 30 positive, and 30 neutral words were selected (see *Chapter 4: Preparatory Study – Word Rating Study*). Words were matched across valence categories for word frequency, word frequency class, word length, imagery, and age-relevance. Positive and negative words were also matched on emotional intensity, that is, the intensity of the valence ratings. In order to standardize words across conditions and lists, three sets of 10 words were created for each valence category. These word sets were built with regard to the word characteristics such that word sets within one valence category did not differ from one another (see Table E1 in the Appendix for means, standard deviations, and results from analyses of variance). Word sets were combined to form three emotion-homogeneous lists (a negative list: 3 x 10 negative words, a positive list: 3 x 10 positive words, and a neutral list: 3 x 10 neutral words) and three emotion-heterogeneous lists (three mixed lists: 10 positive, 10 negative, and 10 neutral words). Table 15 summarizes the characteristics of these six list compositions (three emotion-homogeneous lists and three emotion-heterogeneous lists) with regards to the objective word characteristics and the rating data obtained in the Word Rating Study.

Table 15

Characteristics of the Six Word List Compositions

	Means						Standard Deviations					
	Homogeneous			Heterogeneous			Homogeneous			Heterogeneous		
	N	P	O	M1	M2	M3	N	P	O	M1	M2	M3
Objective Measures												
Frequency	6.18	7.88	9.46	8.66	7.22	7.63	6.87	8.97	10.92	11.18	6.73	9.01
Frequency Class	12.83	12.70	12.33	12.67	12.50	12.70	1.58	1.78	1.81	1.84	1.55	1.80
Length	7.87	8.30	8.03	8.00	8.10	8.10	1.98	1.78	1.71	1.60	2.11	1.77
Word Rating Study												
Valence	2.09	3.94	6.07	4.05	3.99	4.07	0.36	0.66	0.34	1.69	1.79	1.68
Arousal	5.20	3.87	3.01	3.81	4.19	4.08	0.99	1.06	0.94	1.40	1.22	1.39
Control	3.06	3.74	4.82	3.77	3.88	3.97	0.95	1.17	0.58	1.19	1.17	1.20
Imagery	4.68	4.30	4.78	4.71	4.41	4.63	0.78	0.74	0.88	0.76	0.81	0.89
Self-Relevance	2.56	3.79	5.38	3.90	3.94	3.89	0.47	0.96	0.42	1.35	1.39	1.31
Age-Relevance	3.06	3.13	2.98	3.14	3.05	2.98	0.58	0.64	0.69	0.62	0.71	0.57
Derived Score based on Valence												
Emo. Intensity	1.91	0.58	2.07	1.50	1.56	1.50	0.36	0.30	0.34	0.73	0.82	0.70

Note. Emotion-Homogeneous Lists: N = Negative List. P = Positive List. O = Neutral Lists. Emotion-Heterogeneous Lists: M1 = Mixed List 1. M2 = Mixed List 2. M3 = Mixed List 3.

For each list combination, 30 different sequential word orders were generated such that each word appeared at each list position equally often and had each other word equally often as a neighbor (see Figure E1 in the Appendix, in which 30 words were balanced across 30 lists with 30 list positions). This procedure of balancing word orders was based on initial ideas by Tulving (1962) to actually examine clustering effects in the order of the recall output. The balanced word orders were used, on the one hand, to diminish the potential confounding influence of the serial position effect and, on the other hand, to randomize any effects of order on facilitation of subjective clustering.

The serial position effect, that is, that the first items (primacy effect) and the last items (recency effect) are better remembered than items in the middle of the list, has a strong effect on which items get recalled. However, the main focus of this dissertation project was to investigate which words were recalled based on the emotional tone of the words. Thus, list position was a confounding variable to the emotional tone of the words, especially in the emotion-heterogeneous list condition. The balancing procedure provides a means to ensure that each item appeared equally often on each position; or to say it differently, that positive, negative, and neutral words appeared equally often on each position. This does not eliminate the serial position effects but it guarantees that positive, negative, and neutral words have the same probability to participate in this effect.

Another potential confounding factor could be subjective clustering due to the presentation order. Subjective clustering refers to the organization of the to-be-remembered words in the recall output (Tulving, 1962). Which words are recalled together? One source for this organization of which words are recalled together is actually the pattern of which words are presented together. Thus, words that are presented together - spatially or temporally - are also more likely to be recalled together. The balancing procedure ensured that each word has each other word equally often as a neighbor. Thus, it minimizes any effects of presentation order on subjective clustering.

Another advantage of the balancing procedure was that one could easily use the same set of balanced word orders for all sex and age groups. Thus, potential group differences could not be attributed to different word orders used in these groups, which would be possible when word orders were created at random. Moreover, the balancing procedure allowed me to balance words in the emotion-homogeneous and in the emotion-heterogeneous list condition in parallel. This parallel balancing between conditions was on the level of word sets. Thus, for each word set of 10 words, there was a heterogeneous and a homogeneous list, in which the positions of these 10 words were exactly the same.

For the balancing procedure, a theoretical start sequence was generated for an emotion-heterogeneous list of 10 positive, 10 negative, and 10 neutral words. This start sequence functioned as a basis for all other word orders. To generate the start sequence, I constructed all possible triples of negative (N), positive (P), and neutral items (O), that are $3^3 = 27$ triples. All triples were structured to form a consecutive sequence of overlapping triples. For example, the sequence P-N-O-O-P-N consists of the triples P-N-O, N-O-O, O-O-P, O-P-N. One triple, the P-N-O triple appeared twice to cover all 30 words. This procedure of generating a sequence of overlapping triples was done to ensure that all combinations of positive, negative, and neutral words appeared equally often in the lists. Again, this was a means to minimize potential clustering effects in the recall output due to the presentation order. This start sequence is depicted in Figure 6; it is the first list in the heterogeneous list condition.

Based on this start sequence, the other emotion-heterogeneous lists were generated shown in Figure 3. This was done by systematically exchanging positive, negative, and neutral items such that for each list position only one positive, one negative, and one neutral item appeared across all three lists. Then, the words sets were randomly assigned to one list and all words were randomly assigned to one list position. Based on these three heterogeneous lists, the three emotion-homogeneous lists were simply generated by maintaining the position of all words but combining all positive words, all negative words, and all neutral words within one list. This resulted in three emotion-homogeneous lists (see Figure 3).

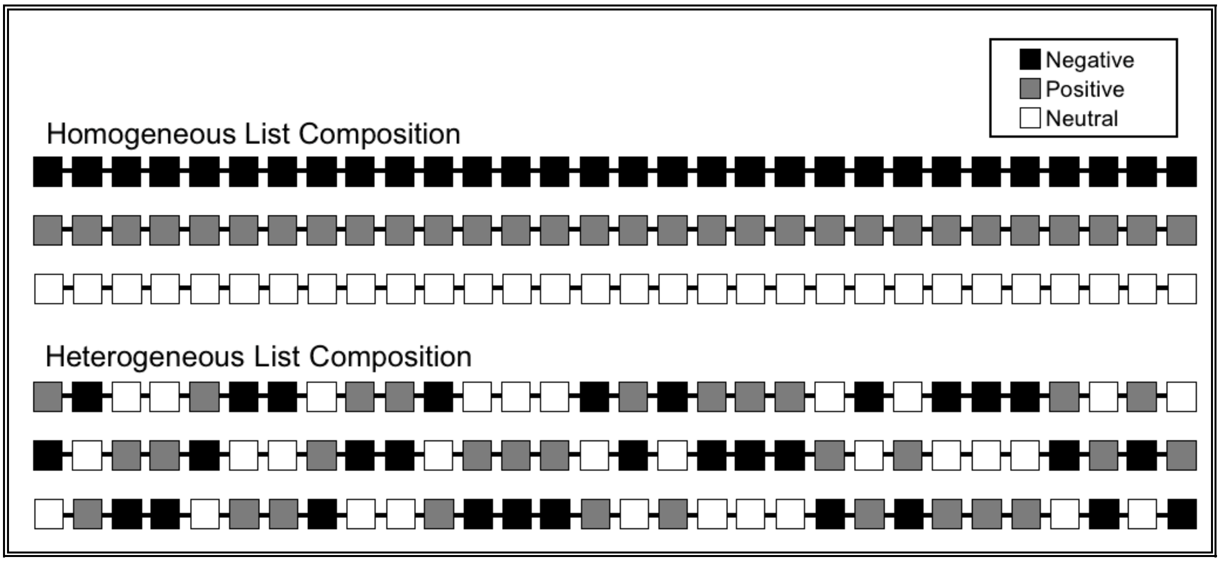

Figure 3 List Composition in the Heterogeneous and Homogeneous List Conditions

Based on each start combination (i.e., the three emotion-heterogeneous and the three emotion-homogeneous start sequences), 30 balanced word orders were created (see Figure E1 in the Appendix). This resulted in 6 lists x 30 orders = 180 lists.

5.1.3 Word Covariates

Similar to the word rating study, I assessed several rating information for the 90 to-be-remembered adjectives. These ratings were used as covariates in follow-up analyses. Words were rated on five dimensions: valence, arousal, young-stereotype, old-stereotype, and self-relevance. In contrast to the word rating study, I used two unipolar scales to assess age-relevance rather than one bipolar scale. This was done to acquire a more differentiated picture on the stereotype-related relevance on these words (see *4.3 Discussion* of the Word Rating Study). For example, it might be the case that one adjective is very relevant for young and for older adults whereas another adjective is not typical for both young and older adults. With the bipolar scale of the Word Rating Study, both adjectives would get a score in the middle. The exact instructions for each dimension are provided in the Appendix B. For each dimension, words were presented in different random orders.

All dimensions were rated on 7-point scales ranging from 1 to 7. For valence, participants were asked to indicate the feeling of pleasantness elicited by each word from *very unpleasant* (1) to *very pleasant* (7). For arousal, participants indicated the feeling of tension elicited by each word from *very relaxed* (1) to *very tensed* (7). For young-stereotype, participants indicated how typical each word is in describing young adults from *not at all typical* (1) to *very typical* (7). For old-stereotype, participants were asked to specify how typical each word is in describing older adults from *not at all typical* (1) to *very typical* (7). For self-relevance, each participant indicated how accurate each word describes himself from *not at all accurate* (1) to *very accurate* (7).

5.1.4 Person Covariates

In order to control for possible relations between interindividual differences in person characteristics and memory for emotionally-toned material (for reviews, see Bower & Forgas, 2001; Eich & Macaulay, 2000), I assessed several indicators for each of four domains: intellectual functioning, affect, well-being, and personality. The domain of cognition was assessed by two measures of fluid intelligence and two measures of crystallized intelligence. The indicators for fluid intelligence were (a) perceptual speed using the Digit Symbol Substitution test (DSS) of the HAWIE-III (Tewes, 1991; a German version of the WAIS-R, Wechsler, 1981), and (b) reasoning using Raven's Standard Progressive Matrices (SPM; Raven, Court, & Raven, 1996; German edition by Heller, Kratzmeier, & Lengfelder, 1998). The two indicators for crystallized intelligence were two vocabulary tests: (a) the Mehrfachwahl-Wortschatz-Intelligenztest-B (MWT-B; Lehrl, 1999), and (b) the Wortschatztest (WST; Schmidt & Metzler, 1992). In addition, Raven's Advanced Progressive Matrices (APM; Raven, 1962; German edition by Kratzmeier, 1980) were used as a filler task in the 1-hour retention interval. However, only few individuals completed all matrices of the APM. Thus, this task was not analyzed.

For affect, three measures were assessed: (a) current mood, (b) positive trait affect, and (c) negative trait affect. Current mood was assessed on a single item, "How do you feel right now?" with a seven-point scale ranging from *very bad* (1) to *very good* (7) with the midpoint *indifferent* (4). Positive and negative trait affect were assessed by the Positive Affect Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988).

The domain of well-being was measured with four indicators: (a) life satisfaction assessed by the Satisfaction with Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985), (b) depression assessed by the Center for Epidemiological Studies – Depression scale (CES-D; Radloff, 1977), (c) subjective physical health as measured by a single item "How good is your physical health at present?", and (d) subjective mental health: "How good is your mental health at present?". Subjective physical health and subjective mental health were assessed on seven-point scales ranging from *very bad* (1) to *very good* (7) with the midpoint *indifferent* (4).

The domain of personality was covered by the assessment of the Big Five factors (i.e., neuroticism, extraversion, agreeableness, conscientiousness, openness). These were measured by the NEO-FFI (Costa & McCrae, 1992; German edition by Borkenau & Ostendorf, 1991).

Single items for affect and well-being were part of the demographic assessment battery in the beginning of the first session. The used standard measure of affect, well-being and personality represent widely-used scales and were assessed in the second session after the 1-week recall. To employ the same response format for all scales and ratings throughout this project, all scales were measured on seven-point scales. One exception was the CES-D scale that was measured on the original four-point response format. The practice of using seven-point scales was done, on the one hand, to minimize the cognitive load for participants by avoiding different response formats, and on the other hand, to improve the psychometric properties (i.e., reliability and validity) of these scales. There is some evidence that the use of seven-point scales (as well as of eight- and nine-point scales) results in enhanced psychometric properties in comparison to the widely-used five-point response format (e.g., Preston & Colman, 2000).[12]

5.1.5 Participants

5.1.5.1 Sample Size and Composition

Seventy-two young adults (ranging in age from 18 to 31 years) and 72 older adults (ranging from 64 to 75 years) were recruited in the local area of Berlin.[13] The sample was stratified by sex and educational level. For the stratification by educational level, participants without a high school degree (German: Hochschulreife or Fachhochschulreife) were defined as low-educated, whereas participants with a high school degree (or above) were defined as high-educated. For each age group and sex, an equal number of low- and high-educated participants were recruited.

Three recruitment strategies were used: (a) advertisements in local newspapers in the city of Berlin, (b) advertisements in local vocational schools in Berlin, and (c) a database of individuals who had participated in previous studies in the Max Planck Institute for Human Development. At recruitment, participants were informed that the present study investigates cognitive tasks in different age groups. For the participants, the study was called *Lernen und Denken* [Learning and Thinking]. Participants received 40 Euro for the two sessions of approximately 2 hours.

5.1.5.2 Socio-Demographic Characteristics

Table 16 gives an overview of socio-demographic characteristics of the total sample and of subsamples of young and older adults. Similar to the word rating study, expected age differences were found in marital status, $\chi^2_{(3)} = 41.89$, $p < .001$, and employment status, $\chi^2_{(8)} = 123.00$, $p < .001$. For marital status, half of the young adults were single (45.8%), whereas the other half was in a long-term partnership (54.2%). The majority of older adults were married or in a long-term partnership (59.7%). However, a significant number of older participants were divorced (13.9%) or widowed (18.1%). For employment status, older adults

[12] The psychometric properties of the affect, well-being, and personality scales are listed in Table E2 in the Appendix. The internal consistencies (i.e., Cronbach's Alpha) were high exceeding generally $\alpha > .70$.

[13] Fourteen participants (four young and ten older adults) were excluded from the study and replaced with newly recruited participants. The criteria for exclusion were: (a) participants did not appear for the second session (three persons), (b) group membership of participants were mixed up (two persons), and (c) participants cheated during the learning phase (nine persons – with one exception, all were older participants).

were predominantly retired (88.9%), whereas young adults were mostly students (56.9%) or unemployed (15.3%).

Regarding the educational level, both subsamples of young and older adults were stratified by high school degree. This procedure of stratification resulted in 36 young and 36 older adults without a high school degree and 36 young and 36 older adults with a high

Table 16

Socio-Demographic Characteristics of the Total Sample and for Subsamples of Young and Older Adults

	Total Sample		Young Adults		Older Adults	
	N = 144		*n* = 72		*n* = 72	
Age (in years)						
Mean (SD)			24.3	(3.5)	69.2	(3.0)
Range			18-31		64-75	
Sex						
Female	72	50.0 %	36	50.0 %	36	50.0 %
Male	72	50.0 %	36	50.0 %	36	50.0 %
Marital Status						
Single	39	27.1 %	33	45.8 %	6	8.3 %
Married, long-term partnership	82	57.0 %	39	54.2 %	43	59.7 %
Divorced	10	6.9 %			10	13.9 %
Widowed	13	9.0 %			13	18.1 %
Education						
Primary education[a]	19	13.2 %	5	6.9 %	14	19.4 %
Lower secondary education[b]	53	36.8 %	31	43.1 %	22	30.6 %
High school[c]	42	29.2 %	31	43.1 %	11	15.3 %
College/University[d]	30	20.8 %	5	6.9 %	25	34.7 %
Years of Education						
Mean (SD)			13.5	(3.6)	13.3	(3.5)
Range			10-21		8-22	
Current Occupation						
Full-time employed	12	8.3 %	9	8.3 %	3	4.2 %
Part-time employed	8	5.6 %	6	8.3 %	2	2.8 %
Unemployed	11	7.6 %	11	15.3 %		
Retired	64	44.4 %			64	88.9 %
Homemaker	2	1.4 %	1	1.4 %	1	1.4 %
Student	41	28.5 %	41	56.9 %		
Job trainee	1	0.7 %	1	1.4 %		
Military Service	1	0.7 %	1	1.4 %		
Freelancer	4	2.8 %	2	2.8 %	2	2.8 %

[a]German: Volks- / Hauptschule. [b]German: Mittlere Reife / Realschulde. [c]German: (Fach-) Abitur. [d]German: Fach- / Hochschulstudium

school degree. Despite this clear match in educational level for young and older adults, the precise distribution revealed substantial age differences in the pattern of educational degrees accomplished, $\chi^2_{(3)} = 28.65$, $p < .001$. Most young adults had either a lower secondary education (43.1%) or a high school education (43.1%), whereas older adults showed a more distributed pattern across all educational levels. Nevertheless, the stratification procedure resulted in a similar number of years of education for young and older adults on average, $F(1,142) = 0.14$, $p = .709$, $\eta_p^2 = .001$. This finding was additional evidence for the success of the stratification procedure in selecting comparable age groups with regard to education.

5.1.5.3 Person Characteristics

Several person covariates were administered with regard to four domains: affect, well-being, personality, and intellectual functioning (see *5.1.4*). For each measure, Table 17 provides means and standard deviations for young and older adults as well as the results of analyses of variance with age (young vs. old) as a between-subjects factor.

As expected from the literature, both age groups differed substantially on many of these scales. However, the pattern was quite complex across scales. Consistent with the literature (e.g., Mroczek, 2001), older adults, in contrast to young adults, reported lower levels of negative trait affect. However, young and older adults did not differ in current mood and positive trait affect. Both age groups stated to be currently and generally in a positive mood.

Older adults showed higher levels of life satisfaction and a tendency ($p < .10$) to report lower values of depression than did young adults. However, age groups differed significantly in their reports of subjective physical health. The majority of older adults reported fair physical health, whereas in the young sample, the majority reported excellent physical health. However, age groups did not differ in their reports of subjective mental health. Both age groups stated high levels of mental health. The finding that older adults reported higher levels of life satisfaction seemed to be somewhat surprising given substantial losses in physical health with advancing age. In the literature on well-being and aging, however, similar findings have often been found and labeled as a *paradox of age-related stability in subjective well-being* (e.g., Diener & Suh, 1997; Mroczek, 2001). Thus, the finding that older adults reported similar or even higher levels of well-being than young adults despite losses in physical health was rather consistent with previous studies.

Table 17

Person Characteristics for Subsamples of Young and Older Adults in the Experiment

	Means		Standard Deviations		ANOVA[a]		
	Young	Old	Young	Old	*F*	*p*	η^2
Affect							
Current Mood [b]	5.19	5.32	1.17	1.20	0.40	.527	.003
Negative Affect (PANAS) [b]	3.20	2.51	1.15	0.81	17.37	<.001	**.110**
Positive Affect (PANAS) [b]	4.99	5.03	1.08	0.87	0.05	.821	<.001
Well-Being							
Life Satisfaction (SWLS) [b]	3.77	4.84	1.34	1.00	29.65	<.001	**.173**
Depression (CES-D) [c]	0.73	0.60	0.41	0.38	3.48	.064	.024
Subjective Physical Health [b]	5.58	4.69	1.24	1.17	19.53	<.001	**.121**
Subjective Mental Health [b]	5.53	5.39	1.14	0.96	0.57	.451	.004
Personality (NEO-FFI)							
Neuroticism [b]	3.63	3.19	1.20	0.87	6.43	.012	**.043**
Extraversion [b]	4.67	4.28	0.84	0.80	8.39	.004	**.056**
Openness [b]	4.92	4.84	0.95	0.81	0.29	.590	.002
Conscientiousness [b]	4.84	5.49	1.08	0.69	18.48	<.001	**.115**
Agreeableness [b]	4.97	5.22	0.83	0.59	4.40	.038	**.030**
Intellectual Functioning							
Vocabulary (WST) [d]	31.56	34.64	4.97	3.71	17.62	<.001	**.111**
Vocabulary (MWT-B) [e]	28.18	32.92	4.88	1.73	60.35	<.001	**.298**
Reasoning (SPM) [f]	50.45	45.29	5.85	6.40	27.73	<.001	**.165**
Perceptual Speed (DSS)	58.54	49.19	11.33	7.71	66.07	<.001	**.321**

Note. Effect sizes in bold were significant at $p < .05$. [a]Degrees of freedom for all *F*-Tests were (1,142). [b]Seven-point scale ranging from 1 to 7. [c]Four-point scale ranging from 0 to 3. [d]Maximum score is 42. [e]Maximum score is 37. [f]Maximum score is 60.

In the personality domain, young adults reported higher levels of neuroticism and extraversion than older adults. However, older adults stated higher values in conscientiousness and agreeableness. And both age groups reported similar levels of openness to new experiences. These age-related differences were consistent with findings from previous studies (e.g., Charles, et al., 2000; McCrae & Costa, 1994; Mroczek & Spiro, 2003; Small, Hertzog, Hultsch, & Dixon, 2003; J. Smith & Baltes, 1999).

As expected from the literature on cognitive aging, the two age cohorts differed in fluid and crystallized intelligence. Young adults performed better than older adults on both indicators of fluid intelligence: reasoning and perceptual speed. In contrast, older adults scored higher than younger adults on both vocabulary tests for crystallized intelligence. This pattern was consistent with past studies investigating age-related differences of cognitive functioning (e.g., Park et al., 2002; Salthouse, 1998; Schaie, 1996).

Taken together, the sample characteristics of young and older adults were consistent with findings obtained in previous studies. Thus, the subsamples of young and older adults were not outstanding or unusual with regards to these person characteristics.[14]

5.1.6 Procedure

This experiment involved two sessions that last for approximately two hours each. During recruitment, the second session was planned to be one week after the first session. For both sessions, participants arrived at the Max Planck Institute for Human Development in small groups of maximal five persons. To ensure privacy, participants were separated by partitions so that they could not look at each other.

Young and older adults were invited separately for both sessions. This was done to accommodate the different presentation times for young and older adults within the learning task (see section *5.1.6.1* for details).

5.1.6.1 Procedure in Session I

Session 1 consisted of four parts: (a) an introduction phase; (b) a learning phase, during which participants learned the adjectives; (c) a 60-minute filler task phase; and (d) a free recall phase, in which participants were asked to recall the adjectives again. At the beginning of Session 1, the experimenter welcomed the participants and gave them a general instruction about the course of the session. Afterwards, participants filled out a short booklet designed to assess demographic characteristics. The experimenter waited until all participants had finished their questionnaires and then introduced the learning task.

[14] Sex-related differences in affect, well-being, personality, and cognitive functioning were considered in the Appendix E. For both age groups, Table E3 provides means and standard deviations separately for men and women. In addition, Table E4 gives the results of analyses of variance with age (young vs. old) and sex (men vs. women) as between-subjects factors. The analyses revealed only three main effects of sex for neuroticism, agreeableness, and reasoning. Women reported higher values in neuroticism (M = 3.64, SD = 1.11) and agreeableness (M = 5.34, SD = 0.65) than men did (neuroticism: M = 3.17, SD = 0.97; agreeableness: M = 4.85, SD = 0.72), whereas men solved somewhat more matrices in the reasoning task (M_{women} = 46.60, SD_{women} = 6.75, M_{men} = 48.91, SD_{men} = 6.47). Moreover, the analyses showed two significant Age x Sex interactions for life satisfaction and openness. For life satisfaction, young women reported higher values (M = 4.01, SD = 1.31) than young men (M = 3.53, SD = 1.34); in contrast older men reported higher values of life satisfaction (M = 5.06, SD = 0.92) than older women (M = 4.63, SD = 1.05). For openness, young men reported somewhat higher values (M = 5.04, SD = 0.83) than young women (M = 4.80, SD = 1.05); in contrast older women reported higher values of openness (M = 5.02, SD = 0.78) than older men (M = 4.67, SD = 0.81). Despite these sex-related differences, men and women seemed to function rather similar than disparate in terms of affect, well-being, personality, and cognitive functioning. The main effects of age were practically unchanged.

In the learning phase, an Apple iBook computer was placed in front of each participant so that participants had a good sight on the screen. Participants were informed that they would see 30 words one after the other on the computer screen and that at the end of the list they should recall as many words as possible. Each participant learned only one list: They were presented this list five times in different orders with subsequent free recall after each trial. This was done to guarantee sufficient acquisition of the material and to investigate possible differences in the learning rates for positive, negative, and neutral material.

The learning task comprised five learning trials, each following the same sequence: (a) a presentation phase, during which one word after another was presented on the computer screen, and (b) a 3-minute free recall phase, during which participants were asked to write down as many words as possible on a single sheet. After this free recall phase, the experimenter collected the recall sheets and asked participants to start the next trial. In this learning phase, no mention was made of a later recall task.

Generally, young adults recall more items and show steeper learning curves than older adults (e.g., Kausler, 1994). In this study, we were not interested in these main effects of age. Instead, our focus was on the relative contribution of positive, negative, and neutral words to total recall in both age groups (i.e., an Age x Valence interaction). To facilitate this comparison, we attempted to ensure approximately similar recall levels in both age groups by means of different presentation rates for young and older adults. For the young adults, each word was displayed for 1000ms on the screen followed by a blank screen for 750ms. For the older adults, each word was displayed for 3000ms and the blank screen appeared for 1000ms. (These values were derived from a pilot study; see Appendix D for details). The different presentation rates should at least reduce mean differences between young and older adults in recall performance. However, the different presentation rates had probably no effect on the steeper learning curves of young adults. Thus, an Age x Trial interaction was expected that young adults recall relative more words with increasing learning trials than older adults.

To minimize interference effects between the learning phase and the 1-hour recall, participants completed cognitive tasks of fluid intelligence during the 1-hour retention interval. These tasks were the DSS and the SPM that were not expected to interfere with memory traces for verbal material or emotionally-laden material. Given large inter-individual differences in speed of completion, some individuals were also asked to complete some or all matrices from Raven's Advanced Progressive Matrices (APM), so that all participants were occupied for approximately one hour.

In the free recall phase, the experimenter handed out a single sheet to each participant. As in the learning phase, participants were asked to write down all words they could still recall. After three minutes, the experimenter collected the written sheets. To ensure that participants actually recalled as many words as possible, the recall task was repeated three times in a row. The reasoning behind this procedure was that the first recall trial may function as an activation of related memory traces. This may lead to an enhanced access to other items on the list. After the recall phase, participants were released from this session.

5.1.6.2 Procedure in Session II

In the beginning of the second session, participants were asked to recall the words again. Similar to the 1-hour recall, this was repeated three times. Afterwards, participants completed three booklets: (a) a booklet covering a collection of questionnaires concerning person characteristics (see section *5.1.4 Person Covariates*), (b) a booklet of both indicators for crystallized intelligence (i.e., MWT-B & WST, see section *5.1.4 Person Covariates*), and (c) a booklet including all rating dimensions for all 90 adjectives (see section *5.1.3 Word Covariates*). Afterwards, participants were paid and released from the experiment.

Table 18 provides a rough overview about the flow of the experiment across both sessions. The main parts were, of course, the learning phase, the 1-hour retention phase, and the 1-week retention phase.

Table 18

Overview of Main Phases in the Procedure of the Experiment

Session	Content	Examples
1	Assessment of Demographic Characteristics	Age, Sex, Education, etc.
	Learning Phase	➔ Learning
	Assessment of Fluid Intelligence	DSS & SPM
	1-Hour Retention Phase	➔ 1-Hour Recall
2	1-Week Retention Phase	➔ 1-Week Recall
	Assessment of Crystallized Intelligence	WST & MWT-B
	Assessment of Personality & Trait Affect	NEO-FFI, etc.
	Assessment of Word Characteristics	Valence, Self-Relevance, etc.

5.1.7 Data Analyses

5.1.7.1 Data Preparation

Data were entered in SPSS 11.5. To check for entry errors, data were entered twice. Discrepant entries were checked and replaced with correct values. All variables were checked for outliers and missing values. Moreover, demographic characteristics (including the indicators for subjective well-being and cognitive fitness) were checked for logical inconsistencies (e.g., reporting an university degree by also reporting only 8 years of schooling in total). Inconsistent or impossible values were replaced with missing values.

For the recall data, all words were entered in the recalled order. Words were coded to be either a correctly recalled word or an intrusion. Spelling errors were generally not considered.

Similar to the word rating study, word characteristics were reorganized to fit the already established database of word ratings. Thus, for each word, means and standard deviations were computed across all participants and across subgroups. Words were matched with the database of adjective to have maximum information available about each word.

5.1.7.2 Statistical Analyses

To analyze the data, different types of analysis of variance were used (including mixed analysis of variance with between- and within-subjects factors). All analyses of variance were done with the MANOVA procedure of SPSS 11.5.

For the main analyses, the dependent variable was the number of recalled words. Depending on the analyses, the typical between-subjects factor was age (young vs. old) and the typical within-subjects factor was trial (e.g., five learning trials). Depending on the condition, whether the analyses were conducted on the heterogeneous or homogeneous lists, valence (negative vs. positive vs. neutral words) was a within-subjects factor in the heterogeneous list condition and a between-subjects factor in the homogeneous list condition.

Follow-up analyses focused on two aspects: person and word characteristics that may influence memory performance for emotionally-toned words. For the person characteristics, several control analyses were conducted with these person characteristics as covariates. For the word characteristics, analyses examined two aspects: (a) the influence of subjectively generated valence categories on memory performance and (b) the influence of word characteristics on recallability (i.e., how easily a word was recalled).

5.2 RESULTS

The result section focused on analyses associated with the central research questions of this dissertation project (see Chapter 3): First, do young and older adults differ in the positive-negative disparity of emotional memory, namely, do older adults show a memory advantage for positively-toned over and above negatively-toned material? Second, are potential differences in general or between age groups driven by selective prioritization processes for negative material? The main analyses comprise three separate sections: (a) analyses of recall performance in the learning phase (see section *5.2.1*), (b) analyses of recall performance in the retention phase (see section *5.2.2*), and (c) analyses of recall of word sets (see section *5.2.6*). These analyses were done to examine the main predictions of this dissertation project: Do older adults favor positive over negative words in memory and more so than young adults supporting the proposed positivity effect in older adults' memory (H1)? Are negative words better recalled than positive and neutral words in an emotion-heterogeneous list context (H2) but not in an emotion-homogeneous list context (H3) documenting emotion-based processing prioritization?

Additional follow-up analyses focused on disentangling the influence of person characteristics (see section *5.2.3*), word characteristics (see section *5.2.7*), and subjective valence categorization (see section *5.2.5*) on the main research predictions. Person and word characteristics were entered as covariates. For subjective valence categorization, the main analyses in the learning phase for the heterogeneous list condition was reran with these subjective valence categories. In addition, the output order of the recalled words was analyzed to examine age-related differences in serial position effects of recalling emotionally-toned material (see section *5.2.4*). If positive information is more salient in older adults memory, older adults might recall positive information prior to negative and neutral information.

5.2.1 Recall Performance in the Learning Phase

Before analyzing the recall performance of young and older adults for positive, negative, and neutral words, an overall mixed analyses of variance was carried out comparing the overall recall performance in the heterogeneous and homogeneous list conditions for young and older adults across the five learning trials. Preliminary analyses on testing the homogeneity of variances across age groups and list conditions revealed equal variances for each learning trial on the univariate level, Trial 1: $F(3,140) = 0.16$, $p = .923$; Trial 2: $F(3,140)$

= 0.53, p = .660; Trial 3: $F(3,140) = 0.10$, p = .961; Trial 4: $F(3,140) = 0.16$, p = .923; Trial 5: $F(3,140) = 0.80$, p = .498, as well as equal variance-covariance matrices on the multivariate level, *Box's M* = 51.39, $F(45,48515) = 1.07$, p = .349, documenting that the multivariate homogeneity assumption for the mixed analysis of variance was not violated.

A 2 x 2 x 5 (Age x List Condition x Trial) mixed analysis of variance was conducted with age (young vs. old) and list condition (heterogeneous vs. homogeneous) as between-subjects factors and trial (five learning trials) as within-subjects factor. The analyses revealed main effects of age, $F(1,140) = 26.94$, $p < .001$, $\eta_p^2 = .161$, and trial, $F(4,560) = 483.68$, $p < .001$, $\eta_p^2 = .776$, and a significant interaction between age and trial, $F(4,560) = 13.66$, $p < .001$, $\eta_p^2 = .089$.

As expected, young adults recalled more words than older adults and all participants recalled an increasing number of words across the five learning trials. This finding is consistent with the cognitive aging literature that both age groups are able to improve their memory performance by giving learning opportunities. The significant interaction between age and trial arose because age differences magnified across trials, that is, young adults showed steeper learning curves. This finding is consistent with expectations about age-related differences in the learning slope. Table 19 provides the overall recall performance and the separate recall performances for the heterogeneous and homogeneous list conditions for all five learning trials for both age groups. The recall levels for young and older adults were in the expected range giving the adjusted presentation times for young and older adults.[15]

Table 19

Overall Recall Performance for Young and Older Adults in the Learning Phase

	Means						Standard Deviations					
	Young Adults			Older Adults			Young Adults			Older Adults		
	M	P	Total	M	P	Total	M	P	Total	M	P	Total
Trial 1	8.53	7.72	8.13	7.03	7.17	7.10	2.40	2.51	2.47	2.65	2.87	2.47
Trial 2	12.67	11.61	12.14	9.86	9.69	9.79	3.23	3.20	3.24	2.77	2.77	2.75
Trial 3	15.19	14.25	14.72	11.89	11.47	11.68	3.76	3.91	3.84	3.53	3.57	3.53
Trial 4	17.72	15.78	16.75	12.92	13.36	13.14	3.95	3.96	4.05	4.60	4.08	4.32
Trial 5	19.19	17.61	18.40	14.58	14.28	14.43	4.83	4.35	4.63	4.18	4.52	4.33

Note. M = Mixed, Heterogeneous List Condition. P = Pure, Homogeneous List Condition.

[15] For both age groups, Table F1 in the Appendix provides descriptive statistics for the recall performance. In particular, this table shows that no participant was able to recall all 30 words demonstrating the absence of a ceiling effect in the recall data. Table F2 provides the intercorrelation matrix for the recall performance in all 11 trials (5 learning trials, 3 1-hour recall trials, & 3 1-week recall trials) separately for young and older adults.

No other effect reached the significance level (see Table F3 in the Appendix for an overview of all effects). Especially the main effect of list condition, $F(1,140) = 1.51$, $p = .221$, $\eta_p^2 = .011$, and the interaction effects for Age x Condition, $F(1,140) = 1.25$, $p = .266$, $\eta_p^2 = .009$; Trial x Condition, $F(4,560) = 0.50$, $p = .736$, $\eta_p^2 = .004$; and Age x Trial x Condition, $F(4,560) = 1.28$, $p = .276$, $\eta_p^2 = .009$, were not significant indicating that the overall recall performance did not differ for the heterogeneous and the homogeneous list condition. This null effect of condition on overall recall performance is relevant in two respects: First, potential differences in the effects of valence in the emotion-heterogeneous and emotion-homogeneous list condition could not be attributed to overall recall differences in both conditions. Second, despite consistent evidence in the memory literature that clustered lists are better remembered than non-clustered lists (e.g., Newman, 1967; Vaughn, 1968), a similar effect of subjective organization on total recall performance is not evident in comparing the heterogeneous (clustered) and the homogeneous (non-clustered) list conditions.

To analyze the recall performance of young and older adults for positive, negative, and neutral words, separate analyses were undertaken for the heterogeneous and homogeneous list conditions. This analysis strategy was used because valence was a within-subjects factor in the heterogeneous list condition and a between-subjects factor in the homogeneous list condition. Thus, a direct comparison on the level of participants was not possible. Table 20 provides mean recall values for positive, negative, and neutral words in the heterogeneous and homogeneous list conditions for both age groups during the learning phase. Initial tests of homogeneity for the different within-subjects levels in the heterogeneous and homogeneous list conditions revealed on the univariate as well as on the multivariate level no violation of the homogeneity assumption (see Table F4 and Table F5 in the Appendix).

5.2.1.1 Heterogeneous List Condition

A 2 x 5 x 3 (Age x Trial x Valence) mixed analysis of variance with age (young vs. old) as a between-subject factor and trial (5 learning trials) and valence (negative vs. positive vs. neutral) as two within-subject factors revealed main effects of age, $F(1,70) = 19.81$, $p < .001$, $\eta_p^2 = .221$; trial, $F(4,67) = 247.98$, $p < .001$, $\eta_p^2 = .780$; and valence, $F(2,69) = 4.75$, $p = .010$, $\eta_p^2 = .064$; and an Age x Trial interaction, $F(4,67) = 9.00$, $p < .01$, $\eta_p^2 = .114$. No other effect reached the significance level. In particular, the expected interaction between age and valence was not significant, $F(2,69) = 1.68$, $p = .190$, $\eta_p^2 = .023$, nor were the Trial x

Valence, $F(8,63) = 0.73$, $p = .665$, $\eta_p^2 = .010$, and the Age x Trial x Valence interactions, $F(8,63) = 0.67$, $p = .721$, $\eta_p^2 = .009$ (see Table F6 in the Appendix for an overview of all effects).[16]

As expected, all participants improved their recall performance across the five learning trials in the emotion-heterogeneous list condition. However, older adults showed flatter learning curves, that is, older adults needed more learning opportunities to reach the same recall level as young adults. Nevertheless, older adults learned new words. These effects are depicted in Figure 4. Separate for emotion-heterogeneous and emotion-homogeneous list conditions, Figure 4 shows for each learning trial the total recall performance of young and older adults aggregated across positive, negative, and neutral words. The depicted values are identical to the values presented in Table 19.

Table 20

Recall Performance in the Learning Phase by List Condition (Heterogeneous and Homogeneous), Age Group, Valence Category (Negative, Positive, Neutral), and Trial

	Means						Standard Deviations					
	Young Adults			Older Adults			Young Adults			Older Adults		
	N	P	O	N	P	O	N	P	O	N	P	O
Heterogeneous List Condition[a]												
Trial 1	3.06	2.64	2.83	2.50	2.42	2.11	1.37	1.07	1.48	2.35	2.23	1.88
Trial 2	4.31	4.03	4.33	3.81	3.03	3.03	1.39	1.44	2.04	2.73	3.45	2.69
Trial 3	5.28	4.69	5.22	4.22	4.06	3.61	1.52	1.77	1.84	3.00	4.34	3.89
Trial 4	6.22	5.75	5.75	4.89	4.31	3.72	1.93	1.86	1.71	4.12	4.34	2.93
Trial 5	6.69	6.25	6.25	5.03	4.97	4.58	1.74	1.84	2.23	5.10	3.80	3.32
Homogeneous List Condition[b]												
Trial 1	7.67	9.42	6.08	7.92	8.00	5.58	1.40	1.38	1.35	2.78	3.02	2.31
Trial 2	11.25	13.42	10.17	10.17	10.17	8.75	1.80	1.30	1.83	2.13	3.49	2.49
Trial 3	13.67	16.08	13.00	12.00	11.75	10.67	1.53	2.00	1.71	3.38	4.12	3.31
Trial 4	15.08	17.58	14.67	13.75	14.00	12.33	1.58	2.08	2.12	4.69	4.31	3.26
Trial 5	16.83	20.08	15.92	14.58	14.42	13.83	1.71	1.80	1.95	5.21	4.64	4.00

Note. [a]Maximum recall within valence categories = 10. [b]Maximum recall within valence categories = 30.
N = Negative words. P = Positive words. O = Neutral words.

[16] The analyses were performed across the three heterogeneous lists. This was done to make the analyses within the heterogeneous list condition comparable to the homogeneous list condition. Including list (the three mixed lists) as an additional between-subjects factor in the mixed analysis of variance added a List x Valence interaction, $F(4,132) = 11.08$, $p < .001$, $\eta_p^2 = .25$. However, the overall pattern of significant effects and their actual size did not change. Therefore, I report only the parsimonious analyses.

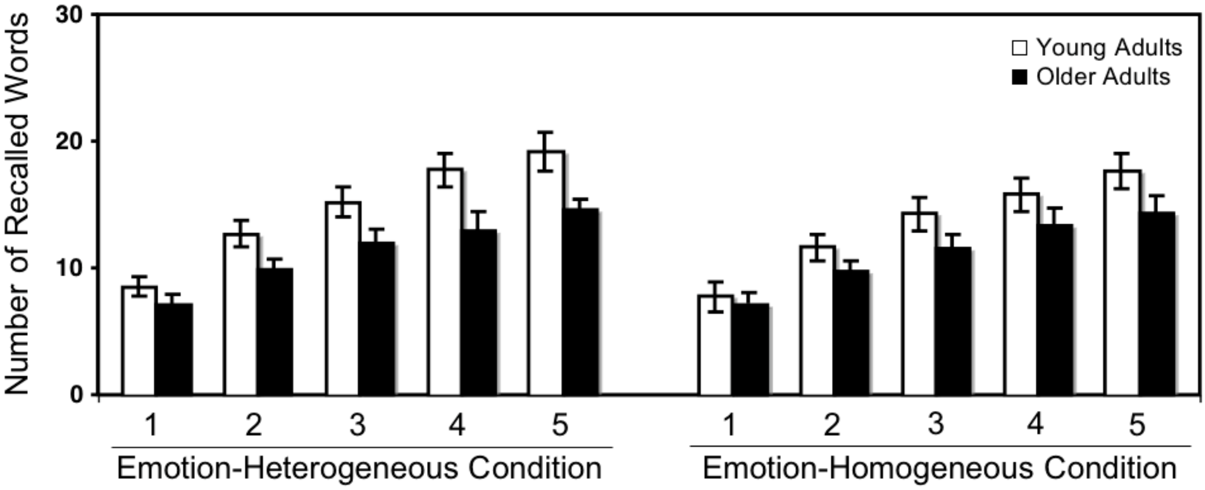

Figure 4 Mean recall as a function of list condition, age group, and learning trial (1-5). Error bars represent 95% confidence intervals.

Follow-up comparisons between valence categories showed that both age groups recalled more negative than positive words, $F(1,70) = 4.82$, $p = .031$, $\eta_p^2 = .031$, and more negative than neutral words, $F(1,70) = 9.39$, $p = .003$, $\eta_p^2 = .118$ (see Table F7 in the Appendix for details of the follow-up comparisons). The comparison between positive and neutral words was not significant, $F(1,70) = 0.21$, $p = .649$, $\eta_p^2 = .003$. These findings are illustrated in Figure 5. Separate for the emotion-heterogeneous and emotion-homogeneous list conditions, Figure 5 presents recall performance for positive, negative, and neutral words aggregated across age groups and learning trials. Thus, in the emotion-heterogeneous list condition, both age groups showed a negativity effect in their recall performance. This finding is consistent with Hypothesis 2 that there is an emotion-based processing prioritization for negatively-toned information when positive and neutral information compete on processing resources.

Despite the main effect of valence, no other interaction effect with valence reached significance. In particular, there was no significant Age x Valence or Age x Trial x Valence interactions indicating that young and older adults did not differ in their recall pattern with regards to positive, negative, and neutral words. This null finding did not support the idea that older adults prioritize positive over negative material in memory and more so than young adults do (see Charles et al., 2003). Thus, the hypothesis for a positivity effect in older adults' memory was not supported (Hypothesis 1).

Moreover, there was no significant Trial x Valence and Age x Trial x Valence interaction documenting that the learning rates for positive, negative, and neutral words did not differ across the five learning trials. Thus, there was no evidence for different learning curves for positive, negative, and neutral words for young and older adults. This finding did not support Hypothesis 4 that age-related differences in remembering emotional information magnify with increasing learning opportunities.

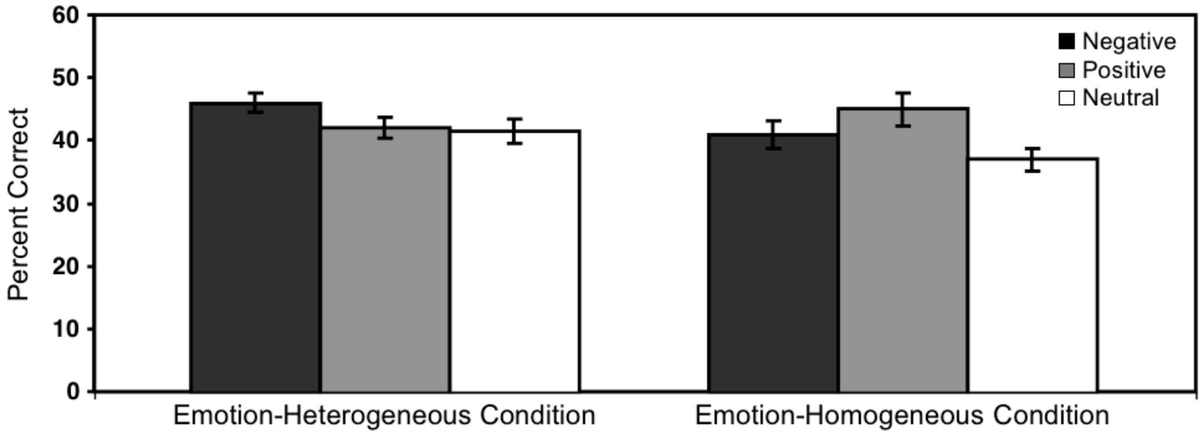

Figure 5 Valence-specific recall as a function of list condition. To facilitate comparison between conditions, scores in the figure represent percent of recalled words within one valence category (negative, positive, or neutral) aggregated across trials and age groups. Error bars represent 95% confidence intervals.

5.2.1.2 Homogeneous List Condition

For the emotion-homogeneous list condition, a 2 x 5 x 3 (Age x Trial x Valence) mixed analysis of variance was conducted, this time with age and valence as between-subject factors and trial as a within-subjects factor. Similar to the heterogeneous list condition, we found main effects for age, $F(1,66) = 8.93$, $p = .004$, $\eta_p^2 = .119$; trial $F(4,67) = 228.19$, $p < .001$, $\eta_p^2 = .776$; and valence, $F(2,66) = 3.52$, $p = .035$, $\eta_p^2 = .096$; and a significant interaction between age and trial, $F(4,63) = 5.64$, $p < .001$, $\eta_p^2 = .079$. There was no other significant effect. In particular, there was no significant Age x Valence, $F(2,66) = 1.01$, $p = .370$, $\eta_p^2 = .030$; Trial x Valence, $F(8,128) = 0.12$, $p = .919$, $\eta_p^2 = .012$; or Age x Trial x Valence interaction, $F(8,128) = 0.48$, $p = .901$, $\eta_p^2 = .014$ (see Table F6 in the Appendix for an overview of all effects).

Again, young adults recalled more words than older adults and both age groups recalled an increasing number of words with increasing learning opportunities. However, young adults showed steeper learning rates than older adults. Thus, age differences magnified with increasing learning trials. This pattern has been consistently found in both conditions.

As illustrated in Figure 5, the main effect of valence revealed a different underlying pattern than found in the heterogeneous list condition. Follow-up comparisons between valence categories showed that both age groups recalled significantly more positive than neutral words, $F(1,44) = 7.32$, $p = .010$, $\eta_p^2 = .143$. The comparisons between negative and positive words, $F(1,44) = 1.54$, $p = .222$, $\eta_p^2 = .034$, as well as between negative and neutral words, $F(1,44) = 1.97$, $p = .168$, $\eta_p^2 = .043$, were not significant (see Table F7 in the Appendix for details). Thus, there was no evidence for a positive-negative disparity of emotional memory in the emotion-homogeneous list condition. This finding is consistent with the hypothesis that no memory differences occur in comparing homogeneous lists of positive and negative words due to the fact that emotion-homogeneous lists do not provide the opportunity to differentially prioritize positive and negative information (Hypothesis 3).

There was no other significant effect. Particularly, there was no evidence for an Age x Valence or Age x Trial x Valence interaction documenting that young and older adults did not recall positive, negative, and neutral words differently. Again, there was no evidence for a Trial x Valence interaction signifying that the learning rates for positive, negative, and neutral lists did not differ.

5.2.1.3 Summary

The analyses on the learning phase were done to examine the main research question of this dissertation project, whether there is a memory advantage of older adults for positively-toned material. The findings were clear-cut. Aside from a main effect of age that young adults recalled more words on average than older adults, the analyses showed no age differences for the positive-negative disparity in remembering emotionally-toned words. Moreover, there were no significant Age x Valence or Age x Valence x Trial interactions within both the emotion-heterogeneous and the emotion-homogeneous list condition. Thus, young and older adults showed highly similar recall pattern with regard to valence. This pattern does not support recent proposals about a positivity effect in older adults memory (e.g., Charles et al., 2003). In both list conditions, no memory advantage for positive over negative words in older adults' memory was evident.

Across learning trials, we found a main effect of trial and an interaction between trial and age for both conditions. Young and older adults improved their recall performance across learning trials. This finding is consistent with expectations that even older adults show learning potential in acquiring new information. However, young adults showed even steeper learning curves than older adults, that is, older adults needed more learning trials to reach the same level of performance than young adults. There was no other significant interaction with trial. Especially, there was no significant interaction between age group, valence and learning trials documenting that the learning rates for positive, negative, and neutral words did not differ in general and between young and older adults. Thus, the recall pattern for positive, negative, and neutral words was quite consistent across trials.

Finally, we found, as expected, a differential pattern of recalling positive and negative material across the two list conditions. This pattern suggested a heterogeneity-homogeneity context distinction. In the context with strong cues for processing prioritization (emotion-heterogeneous context), participants recalled more negative than positive words and more negative than neutral words, whereas in the context with no cues for processing prioritization (emotion-homogeneous context), participants who learned only positive words recalled more than those who learned only neutral words but there was no difference in recall for negative and positive words. This finding is consistent with the idea of emotion-based processing prioritization for negative information when positive and negative information compete on processing resources.

However, in comparing effect sizes, the main effects of age and trial were much larger than the main effect of valence. This was the case within both list conditions. Thus, the found effects of valence were rather small effects. To ensure that the valence effects were robust follow-up analyses were conducted. These follow-up analyses are described in more detail in following sections.

5.2.2 Recall Performance in the Retention Phase

The analyses of retention were guided by three research questions: First, does the recall of emotional and neutral words change with time? Specifically, it was hypothesized that negative information might show stronger decay in memory than positive information (i.e., Hypothesis 8)? Second, does differential forgetting for positive and negative words occur? This research question is instructed by the research prediction that negative information might show stronger decay in memory than positive information (Hypothesis 9). And finally, are forgetting curves for positive and negative words different for young and older adults? In particular, is the proposed positivity effect in older adults’ memory (only) evident in long-term memory?

The analyses in the retention phase were done on aggregated means across the three recall trials for the 1-hour and the 1-week retention intervals.[17] To evaluate the effect of age in learning on retention, three performance measures were included in the analyses: recall at final learning trial (Trial 5), recall after 1-hour, and recall after 1-week retention interval. Similar to the analyses strategy adopted for performance during the learning phase, and before analyzing retention separately for positive, negative, and neutral words, an overall mixed analysis of variance was conducted. Initial tests of univariate and multivariate homogeneity of variances across groups revealed no violation of the homogeneity assumption (univariate tests: Learning Trial 5: $F(3,140) = 0.80$, $p = .498$; 1-Hour Recall: $F(3,140) = 0.48$, $p = .699$; 1-Week Recall: $F(3,140) = 0.86$, $p = .463$; multivariate test: *Box's M* = 9.58, $F(18,69261) = 0.51$, $p = .955$).

The number of recalled words in the retention phase was analyzed with a 2 x 2 x 3 (Age x List Condition x Time) mixed analysis of variance with age (young vs. old) and list condition (heterogeneous vs. homogeneous) as between-subjects factors and time (learning trial 5 vs. 1-hour recall vs. 1-week recall) as within-subjects factor. The analysis revealed main effects of age, $F(1,140) = 30.15$, $p < .001$, $\eta_p^2 = .177$, and time, $F(2,280) = 284.65$, $p < .001$, $\eta_p^2 = .760$, and a significant interaction between age and time, $F(2,280) = 18.34$, $p < .001$, $\eta_p^2 = .116$.

As expected from the literature, both young and older adults showed substantial declines (i.e., forgetting) across the 1-week period. For young adults, recall performance after 1-week had dropped to 59.1% of their recall performance at the end of the learning phase (i.e., learning trial five), whereas for older adults, recall performance dropped to 48.4%. The main effect of age indicated that young adults recalled more words in the retention phase than older adults did. This was not surprising given that young adults acquired more words (M = 18.40) than older adults (M = 14.43) in the learning phase. Thus, the forgetting process for young adults started from a higher level than the forgetting process for older adults. However, the interaction between age and time signified that older adults forgot even more words across the 1-week period than young adults did. Table 21 provides the overall recall performance and the separate recall performances for the heterogeneous and homogeneous list conditions in the retention phase for both age groups.

[17] Table F1 in the Appendix provides descriptive statistics for all trials in the 1-hour and 1-week retention intervals. As indicated in the table, all participants were able to recall at least one word in the 1-hour retention interval. Thus, there was no evidence for potential floor effects in the recall data.

Table 21

Overall Recall Performance for Young and Older Adults in the Retention Phase

	Means						Standard Deviations					
	Young Adults			Older Adults			Young Adults			Older Adults		
	M	P	Total	M	P	Total	M	P	Total	M	P	Total
Trial 5 Recall	19.19	17.61	18.40	14.58	14.28	14.43	4.83	4.35	4.63	4.18	4.52	4.33
1-Hour Recall	15.33	14.80	15.07	10.44	10.00	10.22	5.44	5.29	5.33	4.81	4.65	4.65
1-Week Recall	11.57	10.87	10.87	6.30	7.68	6.99	5.59	5.21	4.81	4.79	4.74	4.78

Note. M = Mixed, Heterogeneous List Condition. P = Pure, Homogeneous List Condition.

No other effect reached the significance level (see Table F8 in the Appendix for an overview of all effects). Especially the main effect of list condition, $F(1,140) = 0.04$, $p = .838$, $\eta_p^2 < .001$, and the interaction effects for Age x Condition, $F(1,140) =1.30$, $p = .256$, $\eta_p^2 = .009$; Time x Condition, $F(2,280) = 0.15$, $p = .859$, $\eta_p^2 = .001$; and Age x Time x Condition, $F(2,280) = 2.10$, $p = .125$, $\eta_p^2 = .015$, were not significant indicating that the overall recall performance in the retention phase did not generally differ for the heterogeneous and the homogeneous list condition. Moreover, the heterogeneous and homogeneous list conditions were not differently affected by the long-term retention interval.

Separate analyses for the heterogeneous and homogeneous list condition were performed to analyze the recall performance for positive, negative, and neutral words in young and older adults. Preliminary tests of homogeneous variances across between-subjects groups revealed on the univariate as well as on the multivariate level no violation of the homogeneity assumption for the analyses of variance. For both retention intervals, Table 22 provides means and standard deviations separately for young and older adults and for the heterogeneous and homogeneous list condition. To facilitate comparisons between the learning phase and the retention phase, scores from learning trial five are reprinted in Table 22. (Table F9 in the Appendix provides mean recall data for all three trials within the 1-hour and 1-week retention interval.)

5.2.2.1 Heterogeneous List Condition

For the heterogeneous list condition, the number of recalled words in the retention phase were subjected to a 2 x 3 x 3 (Age x Time x Valence) mixed analysis of variance with age (young vs. old) as a between-subject factor and time (learning trial 5 vs. 1-hour vs. 1-week) and valence (negative vs. positive vs. neutral) as two within-subject factors. The analysis revealed main effects of age, $F(1,70) = 20.41$, $p < .001$, $\eta_p^2 = .226$; time, $F(2,69) =$

132.88, p < .001, η_p^2 = .655; and valence, $F(2,69)$ = 3.83, p = .024, η_p^2 = .052, and a significant interaction between age and time, $F(2,69)$ = 3.83, p < .001, η_p^2 = .153. The effects related to age and time were the same as reported above. Of special interest here are effects limited to valence.

Table 22

Mean Recall Performance in the Retention Phase by List Condition, Age Group, and Valence

	Means						Standard Deviations					
	Young Adults			Old Adults			Young Adults			Old Adults		
	N	P	O	N	P	O	N	P	O	N	P	O
Heterogeneous List Condition[a]												
Learning Trial 5	6.69	6.25	6.25	5.03	4.97	4.58	1.74	1.84	2.23	5.10	3.80	3.32
1-Hour Recall	5.62	4.91	4.81	3.56	3.21	3.22	2.01	1.99	2.32	2.15	1.72	2.06
1-Week Recall	4.13	3.91	3.53	2.25	2.13	1.92	2.44	1.81	2.11	2.06	1.76	1.77
Homogeneous List Condition[b]												
Learning Trial 5	16.83	20.08	15.92	14.58	14.42	13.83	1.71	1.80	1.95	5.21	4.64	4.00
1-Hour Recall	14.14	17.72	12.53	10.64	11.81	8.86	5.26	5.47	4.00	5.48	4.61	3.57
1-Week Recall	9.17	13.36	7.97	8.50	8.47	6.06	3.32	4.75	4.68	5.27	5.16	3.56

Note. [a]Maximum recall within valence categories = 10. [b]Maximum recall within valence categories = 30. N = Negative words. P = Positive words. O = Neutral words.

Follow-up comparisons between valence categories showed that both age groups recalled significantly more negative than neutral words, $F(1,70)$ = 8.71, p = .004, η_p^2 = .111. There was a non-significant trend that participants recalled more negative than positive words in the retention phase, $F(1,70)$ = 3.46, p = .07, η_p^2 = .047. The comparison between positive and neutral words was not significant, $F(1,70)$ = 0.37, p = .55, η_p^2 = .005 (see Table F11 in the Appendix for details). Figure 6 depicts means for the number of recalled words in learning trial five, in the 1-hour recall, and the 1-week recall in the right part of the figure. In addition, to help to interpret the magnitude of forgetting in the different valence categories, the left part of Figure 6 provides percent of recalled words based on the number of acquired words in the learning phase. Thus, the recall performance on learning trial five functioned as 100% baseline for the 1-hour and 1-week recalls. This was done separately for young and older adults and for positive, negative, and neutral words.

There were no other significant effects (see Table F10 in the Appendix). Of special interest, there were no significant effects for the interactions of Age x Valence, $F(2,69)$ = 0.66, p = .521, η_p^2 = .009; Time x Valence, $F(4,67)$ = 1.71, p = .149, η_p^2 = .024; and Age x Time x Valence, $F(4,67)$ = 1.17, p = .325, η_p^2 = .016. Contrary to Hypothesis 7, there was no evidence that relative more emotional than neutral words were in the retention phase than in the learning phase. Thus, the enhancement of memory by emotion did not increase with increasing retention interval. There was also no empirical support for Hypothesis 8. Negative words showed no stronger decay than positive words (i.e., Time x Valence interaction). Moreover, the proposed positivity effect in older adults' memory was also not evident in long-term memory for emotionally-toned words.

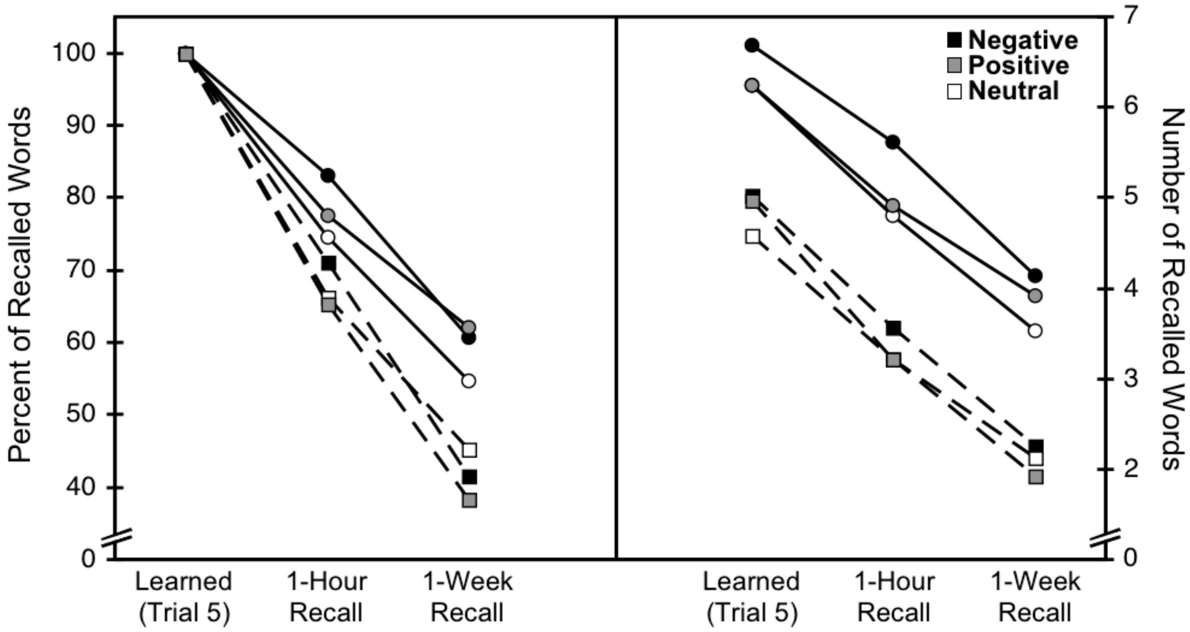

Figure 6 Mean recall for the emotion-heterogeneous list condition for young (○) and older adults (□) in the retention phase as a function of valence. In the figure, the left panel illustrates values representing the relative percentages based on the number of recalled words in learning trial 5, whereas the right panel depicts values for the absolute number of recalled words.

5.2.2.2 Homogeneous List Condition

In the homogeneous list condition, a 2 x 3 x 3 (Age x Time x Valence) mixed analysis of variance with age (young vs. old) and valence (negative vs. positive vs. neutral) as between-subjects factors and time (learning trial 5 vs. 1-hour vs. 1-week) as a within-subjects factor was done on the number of recalled words. The analyses yielded significant main effects of age, $F(1,66) = 11.32$, $p = .001$, $\eta_p^2 = .146$; time, $F(2,65) = 161.67$, $p < .001$, $\eta_p^2 = .710$; and valence, $F(2,69) = 4.60$, $p = .013$, $\eta_p^2 = .122$; as well as significant interactions between age and time, $F(2,65) = 8.07$, $p < .001$, $\eta_p^2 = .109$, and between time and valence,

$F(4,132) = 3.47$, $p = .010$, $\eta_p^2 = .095$. Again, the focus in these analyses was on the two significant effects associated with valence.

Follow-up comparisons between valence categories showed that participants who learned only positive words recalled significantly more words in the retention phase than participants who learned only neutral words, $F(1,44) = 9.67$, $p = .003$, $\eta_p^2 = .180$. This effect was accompanied by an interaction between time and valence, $F(2,43) = 7.53$, $p = .001$, $\eta_p^2 = .146$, signifying that participants who learned only neutral words failed to retrieve these words in a greater extent than participants who learned only positive words. The comparisons between participants who learned negative or neutral words, $F(1,44) = 1.90$, $p = .175$, $\eta_p^2 = .041$, and between participants who learned negative or positive words, $F(1,44) = 2.55$, $p = .118$, $\eta_p^2 = .054$, were not significant (see Table F11 in the Appendix for details). Figure 7 depicts means and percentages for the recall performance in the emotion-heterogeneous list condition of young and older adults for negative, positive, and neutral words. The right part of Figure 7 presents the number of recalled words in the learning trial 5, the 1-hour recall, and the 1-week recall, whereas the left part depicts percentages of recalled words in the 1-hour and 1-week recall based on the recall performance in learning trial five (i.e., 100% baseline).

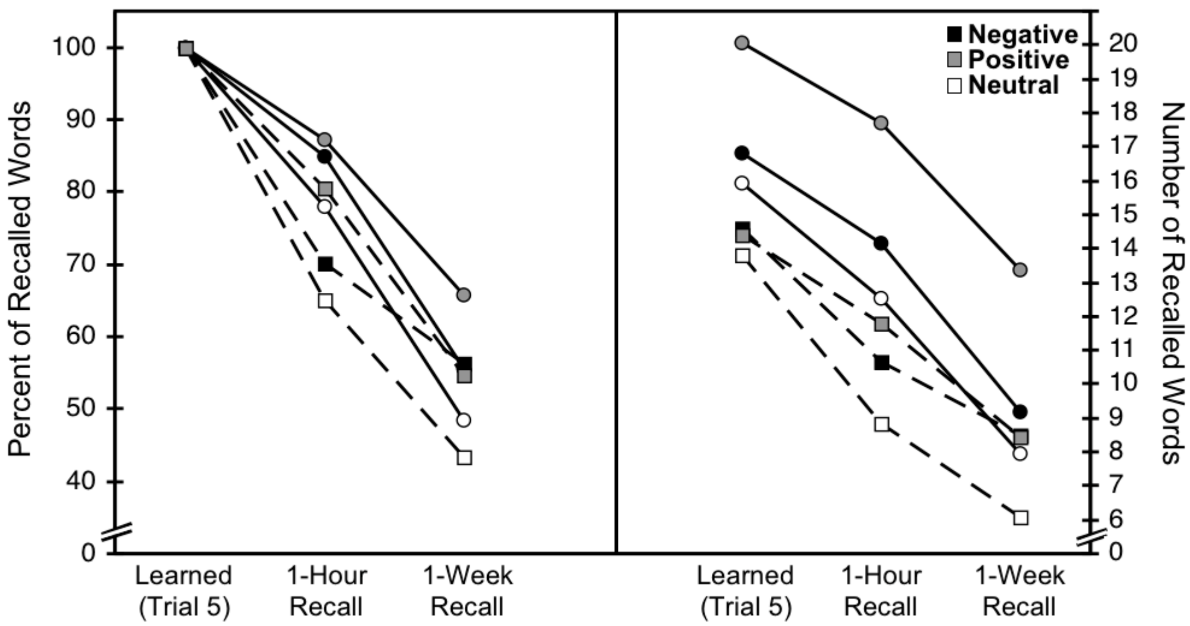

Figure 7 Mean recall for the emotion-homogeneous list condition for young (○) and older adults (□) in the retention phase as a function of valence. In the figure, the left panel illustrates values representing the percentage of recalled words as a function of the recalled words in trial 5, whereas the right panel depicts values for the number of recalled words.

There was no other significant effect (see Table F10 in the Appendix). In particular, there was no significant effects for the interactions of Age x Valence, $F(2,66) = 1.03$, $p = .364$, $\eta_p^2 = .030$, and Age x Time x Valence, $F(4,132) = 0.62$, $p = .648$, $\eta_p^2 = .018$. Thus, there was no evidence for age-related differences in remembering positive, negative, and neutral words in the retention phase. Specifically, there was no evidence for the proposed positivity effect in older adults' memory (see Hypothesis 1).

The significant interaction between time and valence for positive and neutral words provided some support for the research prediction that emotionally-toned material show greater long-term consolidation than neutral material (Hypothesis 8). Although not significant, mean values are suggestive of a tendency that young adults forgot negative words easier than positive words across the 1-week period, whereas older adults showed no differences in forgetting positive and negative words. This pattern of findings would be consistent with Hypothesis 8 – at least for the young adults. Nevertheless, the pattern was not consistent with the proposed positivity effect in older adults' memory. Moreover, the findings provided no support for the idea that older adults show a positivity effect in long-term memory.

5.2.2.3 Summary

The analyses in the retention phase were guided by three research questions: First, is the memory enhancement by emotion greater with increasing retention interval? Second, do participants showed differential forgetting for positive and negative words; in particular, do negatively-toned words decayed faster than positively-toned words and does this depend on list context? Finally, is the proposed positivity effect in older adults' memory evident in long-term memory (rather than in short-term learning)?

In the emotion-heterogeneous list condition, the analyses revealed no evidence for different forgetting curves of positive, negative, and neutral words. The recall pattern in the retention phase resembled the pattern in the learning phase. Negative words were better remembered than neutral words. However, the significant comparison between negative and positive words in the learning phase was only a trend in the retention phase. Nevertheless, the effect size for this comparison in the retention phase ($\eta_p^2 = .047$) was in the same range as in the learning phase ($\eta_p^2 = .031$) indicating a consistent recall pattern across the learning and retention phases.

In addition, there was no evidence for age-related differences in forgetting emotionally-toned words in the emotion-heterogeneous list condition, especially there was no support for an advantage of positively-toned words in older adults' memory over a period of one week. Moreover, inspecting means, the data would suggest that young adults forgot neutral words much easier than negative and positive words whereas older adults forgot positive words much easier than negative and neutral words. These mean differences were not significant; however, the mean differences in the emotion-heterogeneous list condition were not in the direction to support a tendency for a positivity effect in older adults' memory.

In the emotion-homogeneous list condition, the analyses showed some evidence for differential forgetting. Participants who learned only neutral words showed greater declines across the 1-week period than participants who learned only positive words. Thus, in the emotion-homogeneous list condition, neutral words decayed faster than positive words. One explanation for this effect might be that the 30 neutral words were more loosely connected in semantic space than the 30 positive words. Thus, the retrieval of a neutral word did not provide as many cues as the retrieval of a positive word in retrieving other neutral or positive words. To say it differently, the semantic network of positive words (and of negative words?) has a higher density than the semantic network of neutral words resulting in differences in retrieval. The rank order of the level of recall performance was participants who learned only neutral words, who learned only negative words, and who learned only positive words in ascending order. The recall performance for the negative-homogeneous group was not significantly different from the positive-homogeneous and neutral-homogeneous groups. This pattern of findings in the retention phase was highly similar to the pattern in the learning phase.

Regarding age-related differences in the emotion-homogeneous list condition, there was no evidence for an advantage of positively-toned material in older adults long-term memory. Inspecting means, the recall pattern in the retention phase would suggest that young adults remember positively-toned words better than negatively-toned words whereas older adults showed a similar recall pattern for positively- and negatively-toned words. Again, these means were not significantly different but confirmed that there was actually no tendency in the data for a positivity effect in older adults' memory.

In sum, there was only partial evidence for different forgetting curves for positive, negative and neutral words. Neutral words were more easily forgotten only in the emotion-homogeneous list condition. There was no evidence for a positive-negative disparity in forgetting. This finding does not support the expectation that memory for negative words might show stronger forgetting than memory for positive words. In addition, there was no evidence for age-related differences in remembering emotionally-toned words across a 1-

week period. In particular, there was no evidence for an advantage of older adults' memory for positively-toned words in both list conditions. Thus, this finding does not support speculations that the proposed positivity effect might be only evident in long-term memory.

5.2.3 Follow-up Analyses on Person Characteristics

To test the robustness of the effects and to investigate the possible influence of person characteristics on the obtained findings, I conducted analyses of covariance (ANCOVAs) for the measured person characteristic including measures of affect, well-being, personality, and cognition. The analysis of covariance has two major advantages: First, it provides a way to statistically equate groups that have somewhat different values on the covariate. Second, covariates can reduce unexplained variance namely differences between age-cohort groups or experimental groups, thereby increasing power to detect an effect of the independent variables. Additionally, not adjusting for a covariate may lead to wrong conclusions. Specifically, the random assignment in the experiment did not guarantee that groups were the same. Thus, the analysis of covariance provides a mean to address these queries.

The analyses were guided by the questions of whether interindividual differences in person characteristics may underlie found differences between valence categories or whether interindividual differences in person characteristics may actually hide 'real' age-related differences in remembering positive, negative, and neutral words. Before describing the analyses of covariance, I provide information about the zero-order correlations between person characteristics and overall recall performance. These correlations are informative with regards to the general magnitude of associations between person characteristics and recall performance.

5.2.3.1 Correlations between Person Characteristics and Overall Recall

Separately for young and older adults, Table 23 presents zero-order correlations between person characteristics and overall recall performance. For the recall performance, Table 23 provides information for all five learning trials and the means of the 1-hour and the 1-week recall trials.[18]

The correlation pattern was unambiguous. For both age groups, overall recall performance was primarily associated to indicators of intellectual functioning. For young adults, the correlations between overall recall performance and measures of fluid intelligence

[18] Table F12 in the Appendix provides the correlation matrix between all person characteristics. Moreover, Table F2 provides the correlation matrix between all recall trials.

(i.e., reasoning and speed) were somewhat larger (.18 ≤ *r* ≤ .42) than correlations with measures of crystallized intelligence (i.e., vocabulary scores including years of education, .15 ≤ *r* ≤ .30). For the older adults, the correlation pattern was quite the opposite: The correlations to overall recall performance were somewhat stronger for measures of crystallized intelligence (including years of education, .29 ≤ *r* ≤ .50) than for measures of fluid intelligence (.19 ≤ *r* ≤ .43).

Aside from indicators of intellectual functioning, three person characteristics showed significant correlations to overall recall performance in one or two trials. For young adults, these person characteristics were openness and agreeableness; for older adults, these person characteristics were life satisfaction and openness. However, only openness in older adults was consistently associated with overall recall performance (.34 ≤ *r* ≤ .53). This finding might indicate that older adults who maintained the curiosity and openness to make new experiences could also maintain their cognitive performance to some degree (e.g., Lövdén, Ghisletta, & Lindenberger, 2005). Indicators of affect were generally not related to overall recall performance (-.11 ≤ *r* ≤ .14).

Table 23

Correlations between Person Characteristics and Recall Performance in the Learning Phase (Trials 1-5) and Retention Phase (1-Hour and 1-Week Recall) for Young and Older Adults

	Young Adults							Older Adults						
	1	2	3	4	5	H	W	1	2	3	4	5	H	W
Affect														
Current Mood	.03	-.01	.01	.02	.05	-.01	.02	.09	-.01	.07	.14	.02	.11	.04
Negative Affect (PANAS)	-.09	-.04	.00	-.04	.03	-.01	.08	-.11	.01	-.08	-.07	.00	-.07	-.07
Positive Affect (PANAS)	.04	.13	.13	.05	.03	.01	.05	.06	-.02	.07	.06	.06	.05	.02
Well-Being														
Life Satisfaction (SWLS)	.13	.10	.14	.05	.06	.09	.19	-.01	-.23	-.17	-.21	**-.26**	**-.25**	-.22
Depression (CES-D)	-.16	-.11	-.03	-.07	-.06	-.02	.06	-.03	.13	.04	.04	.09	-.04	.10
Subjective Physical Health	-.04	.03	.11	.01	.06	.08	.04	.06	-.18	.01	.10	-.01	.06	-.01
Subjective Mental Health	.01	.07	.06	.09	.06	.12	-.02	.16	.04	.12	.18	.07	.16	.19
Personality (NEO-FFI)														
Neuroticism	-.09	-.09	-.06	-.03	.00	-.05	.04	-.16	-.01	-.07	-.10	-.03	-.13	-.14
Extraversion	.12	.18	.10	.03	.01	.00	-.01	.04	.01	.15	.07	.02	.12	.12
Openness	**.24**	.21	.20	.14	.10	.11	**.25**	**.34**	**.46**	**.53**	**.40**	**.41**	**.44**	**.38**
Conscientiousness	-.07	-.12	-.19	-.10	-.08	-.06	-.02	.01	-.04	-.07	-.07	-.08	-.15	-.13
Agreeableness	.13	**.23**	.20	.17	.09	.18	.15	-.04	.00	.11	.03	.09	.09	.06
Intellectual Functioning														
Vocabulary (WST)	**.27**	**.27**	**.30**	**.25**	.15	.20	.19	**.35**	**.45**	**.50**	**.48**	**.46**	**.49**	**.34**
Vocabulary (MWT-B)	**.27**	.23	.21	.23	.22	**.25**	.20	**.40**	**.44**	**.39**	**.38**	**.35**	**.42**	**.29**
Reasoning (SPM)	.23	.18	**.29**	**.33**	**.32**	**.26**	.21	**.29**	.19	.23	**.26**	**.30**	**.43**	**.39**
Perceptual Speed (DSS)	**.34**	**.38**	**.40**	**.42**	**.33**	**.37**	**.36**	**.32**	**.27**	**.25**	**.36**	**.26**	.20	.21
Years of Education	.12	**.24**	**.27**	**.27**	.18	.14	.17	**.33**	**.36**	**.39**	**.41**	**.38**	**.45**	**.37**

Note. Values in bold are significant at $p < .05$. 1 - 5 = Learning Trials 1 – 5. H = 1-Hour Recall. W = 1-Week Recall.

5.2.3.2 Person Characteristics as Covariates

In order to test whether findings in the emotion-heterogeneous and emotion-homogeneous list conditions could be explained by interindividual differences in well-being, affect, personality, and intellectual functioning, I reran the main analyses for the learning phase with these person characteristics as covariates. In the analyses, covariates were entered individually resulting in 34 (2 list conditions x 17 person characteristics) analyses of covariance.

Before describing the findings, I would like to mention a peculiarity of analyses of covariance in designs with one or more within-subjects factors. In within-subjects or mixed designs, there are two distinct classes of covariates. One is the varying covariate, in which the participant has one covariate for each level of the within-subjects factor. The second class is the constant covariate, in which each participant has a covariate that applies to that participant (rather than to each level of the within-subjects factor). In the present experiment, all person characteristics were constant covariates that were measured only once. The peculiarity is that constant covariates could mathematically not affect the impact of within-subjects effects. Thus, entering constant covariates could not change the effects of within-subjects factors. Constant covariates could only affect the influence of between-subjects factors. That said the covariates could actually not affect the main effect of valence in the emotion-heterogeneous list condition. However, including constant covariates can alter the main effect of valence in the emotion-homogeneous list condition.

Table 24 provides the results of the analyses of variance separately for the emotion-heterogeneous and emotion-homogeneous list condition. Only effect sizes are depicted to facilitate the comparison between different analyses (and to reduce visual complexity of the table). The effect sizes for the analyses without a covariate are shown in the top of Table 24 (c.f., section *5.2.1 Recall Performance in the Learning Phase*).

As expected from the comments about constant covariates, the analyses of covariance for the emotion-heterogeneous list condition revealed practically no changes for the main effects of trial and valence.[19] Moreover, the interactions between age and valence as well as between age, valence, and trial did not change at all. Only the main effect of age showed some alterations by including covariates. In particular, the main effect of age disappeared by adjusting for perceptual speed. This finding is consistent with the idea that cognitive slowing

[19] Small deviations from the exact effect sizes were due to the exclusion of one young participant who obviously produced incorrect test scores for the cognitive tests.

(i.e., speed of processing) is the primary variable in explaining age-related declines in cognitive functioning (Salthouse, 1991, 1996).

Table 24

Changes in Effect Sizes by Entering Covariates into the Main Analyses of the Learning Phase

	Heterogeneous List Condition							Homogeneous List Condition						
	A	T	V	AT	AV	TV	ATV	A	T	V	AT	AV	TV	ATV
Without Covariates	**.22**	**.78**	**.06**	**.11**	.02	.01	.01	**.12**	**.78**	**.10**	**.08**	.03	.01	.01
Affect														
Current Mood	**.23**	**.78**	**.06**	**.11**	.02	.01	.01	**.12**	**.78**	**.10**	**.08**	.03	.01	.01
Negative Affect (PANAS)	**.16**	**.78**	**.06**	**.11**	.02	.01	.01	**.16**	**.77**	.09	**.08**	.04	.01	.01
Positive Affect (PANAS)	**.24**	**.78**	**.06**	**.11**	.02	.01	.01	**.12**	**.77**	**.09**	**.08**	.03	.01	.01
Well-Being														
Life Satisfaction (SWLS)	**.18**	**.78**	**.06**	**.11**	.02	.01	.01	**.10**	**.78**	**.10**	**.08**	.03	.01	.01
Depression (CES-D)	**.22**	**.78**	**.06**	**.11**	.02	.01	.01	**.12**	**.78**	**.10**	**.08**	.03	.01	.01
Subjective Physical Health	**.20**	**.78**	**.06**	**.11**	.02	.01	.01	**.09**	**.78**	**.10**	**.08**	.03	.01	.01
Subjective Mental Health	**.24**	**.78**	**.07**	**.13**	.03	.01	.01	**.08**	.78	**.11**	**.08**	.05	.01	.01
Personality (NEO-FFI)														
Neuroticism	**.22**	**.78**	**.06**	**.11**	.02	.01	.01	**.13**	**.78**	**.10**	**.08**	.04	.01	.01
Extraversion	**.20**	**.78**	**.06**	**.11**	.02	.01	.01	**.14**	**.78**	**.10**	**.08**	.04	.01	.01
Openness	**.23**	**.78**	**.06**	**.11**	.02	.01	.01	**.12**	**.78**	.08	**.08**	.03	.01	.01
Conscientiousness	**.19**	**.78**	**.06**	**.11**	.02	.01	.01	**.09**	**.78**	**.11**	**.08**	.04	.01	.01
Agreeableness	**.25**	**.78**	**.06**	**.11**	.02	.01	.01	**.12**	**.78**	**.10**	**.08**	.03	.01	.01
Cognitive Factors														
Vocabulary (WST)	**.33**	**.79**	**.06**	**.12**	.03	.01	.01	**.21**	**.78**	**.10**	**.08**	.05	.01	.01
Vocabulary (MWT-B)	**.30**	**.78**	**.06**	**.11**	.02	.01	.01	**.19**	**.78**	**.10**	**.08**	.05	.01	.01
Reasoning (SPM)	**.12**	**.78**	**.07**	**.11**	.03	.01	.01	.06	**.78**	**.10**	**.08**	.03	.01	.01
Perceptual Speed (DSS)	.03	**.79**	**.06**	**.10**	.02	.01	.01	.02	**.78**	**.12**	**.08**	.02	.01	.01
Years of Education	**.24**	**.78**	**.06**	**.11**	.02	.01	.01	**.13**	**.78**	**.11**	**.08**	.05	.01	.01

Note. Depicted are main effects of age (A), trial (T), and valence (V), as well as the interactions Age x Trial (AT), Age x Valence (AV), Trial x Valence (TV), and Age x Trial x Valence (ATV). Effect sizes are parital eta-squares (η_p^2). Values in bold are significant at $p < .05$.

For the emotion-homogeneous list condition, the analyses of covariance revealed a robust main effect of valence. The effect sizes were only slightly affected. However, two analyses indicated a non-significant main effect of valence after adjusting for one covariate. These covariates were negative affect, $F(2,64) = 2.98$, $p = .058$, $\eta_p^2 = .085$, and openness, $F(2,65) = 2.96$, $p = .059$, $\eta_p^2 = .084$. Nevertheless, the follow-up comparisons between participants, who learned only positive words, and participants, who learned only neutral words, were still significant. This was the case for negative trait affect, $F(1,42) = 6.24$, $p = .016$, $\eta_p^2 = .129$, and openness, $F(1,43) = 5.81$, $p = .020$, $\eta_p^2 = .119$. Thus, although the main effect of valence dropped somewhat below the alpha level of .05, the follow-up comparisons between positive and neutral words remained significant. Again, the main effect of age demonstrated the greatest changes by including covariates. Specifically, the main effect of age was non-significant after adjustment for interindividual differences in reasoning and perceptual speed. All other effects were basically unchanged by including person characteristics as covariates. In particular, the interactions between Age x Valence and Age x Valence x Trial did not reach the significance level.

5.2.3.3 Summary

Taken together, the analyses of covariance provided systematic evidence that the impact of interindividual differences in person characteristics on the obtained findings in the emotion-heterogeneous and emotion-homogeneous list conditions was small if not negligible. Person characteristics did not consistently affect the main effects of time and valence. Especially the different recall patterns for positive, negative, and neutral words in the emotion-heterogeneous and emotion-homogeneous list condition were unaffected. This is of special interest in light of previous findings in the literature of affect and memory: Dispositions and states in affect-related characteristics are generally thought to be associated to memory performance for emotionally-toned material (for reviews, see Bower & Forgas, 2001; Eich & Macaulay, 2000). In the present study, the analyses of covariance provided no strong evidence for such an association. Regarding the three affect measures, only negative trait affect had a minor impact on the findings for the valence effects in the homogeneous list condition resulting in a reduction of $\Delta\eta_p^2 = .014$ for the comparison of positive and neutral words. Thus, negative trait affect explained approximately 1% of the variance. This is quite a small effect. Moreover, the state measure of current mood showed no impact at all on memory performance for emotionally-toned material. This is somewhat surprising given that current mood was measured approximately 10 minutes prior to the learning task, that is, temporally very close to the memory data.

One has to admit, however, that the research on the relation between affect-related person characteristics and memory performance has used for the most part emotion-induction procedures to elicit higher levels of affect (see Bower & Forgas, 2001). A relation between affect and memory performance might be more evident with strong emotions rather than day-to-day mood levels. In this regards, the findings in the present dissertation project do not provide a strong case against such a relation. However, if such a relation exists, it does not seem to have a major impact on the recall pattern in the present experiment.

The analyses of covariance revealed as well no evidence for a 'hidden' interaction between age and valence in remembering emotionally-toned words. In neither the emotion-heterogeneous nor emotion-homogeneous list condition, the inclusion of covariates did reveal any significant effect in addition to the already established effects.

5.2.4 Follow-up Analyses on Output Order of Recall

The main analyses on the recall data in the emotion-heterogeneous and emotion-homogeneous list condition did not reveal the proposed positivity effect in older adults' memory (e.g., Charles et al., 2003). This was the case in the learning phase (see section *5.2.1*) and in the retention phase (see section *5.2.2*). Moreover, the proposed positivity effect was also not mediated by interindividual differences in person characteristics (see *5.2.3*).

One might wonder whether the proposed positivity effect is evident in the output order of the recalled words. Thus, if positive information is more salient than negative information in older adults' memory, it might well be that older adults recalled the positive words prior to negative and neutral words. This would indicate a prioritization of positive information in the process of retrieval rather than a general memory advantage for positive information.

5.2.4.1 Analyses for the First Three Words Recalled

To analyze the output order of the recall data, I conducted one analysis on the first three words recalled in the emotion-heterogeneous list condition.[20] Table 25 provides means for the number of negative, positive, and neutral words correctly recalled within the first three words. To simplify the analysis, the means were aggregated for the learning phase, the 1-hour recall and the 1-week recall.[21]

A 2 x 3 x 3 (Age x Time x Valence) mixed analysis of variance with age (young vs. old) as a between-subject factor and time (learning phase vs. 1-hour vs. 1-week) and valence (negative vs. positive vs. neutral) as two within-subjects factors revealed significant main effects of age, $F(1,70) = 4.44$, $p = .039$, $\eta_p^2 = .060$, and time, $F(2,140) = 38.87$, $p < .001$, $\eta_p^2 = .357$; and a significant Age x Time interaction, $F(2,140) = 4.96$, $p = .008$, $\eta_p^2 = .066$. No other effect reached the significance level. In particular, the main effect of valence was not significant, $F(2,140) = 0.86$, $p = .425$, $\eta_p^2 = .012$, nor were the interactions between age and valence, $F(2,140) = 0.15$, $p = .883$, $\eta_p^2 = .002$; trial and valence, $F(4,280) = 0.08$, $p = .989$, $\eta_p^2 = .001$; and the Age x Trial x Valence interactions, $F(4,280) = 1.32$, $p = .263$, $\eta_p^2 = .019$

[20] In the emotion-homogeneous list condition, the output order of positive, negative, and neutral words cannot be analyzed due to the fact that all words are from one valence category.
[21] Analyses on the level of trials revealed the same effects.

Table 25

Mean Recall of Negative (N), Positive (P), and Neutral Words (O) in the First Three Recalled Items of Young and Older Adults in the Emotion-Heterogeneous Condition

	Means						Standard Deviations					
	Young Adults			Older Adults			Young Adults			Older Adults		
	N	P	O	N	P	O	N	P	O	N	P	O
Learning Phase	0.95	0.95	1.03	1.09	0.97	0.88	0.35	0.39	0.42	0.34	0.42	0.40
1-Hour Recall	0.94	0.92	0.80	0.87	0.74	0.87	0.55	0.61	0.53	0.59	0.60	0.68
1-Week Recall	0.95	0.80	0.75	0.69	0.61	0.73	0.69	0.62	0.51	0.61	0.62	0.67

Note. The number of negative, positive, and neutral words recalled does not add to 3. This was due to three reasons: (a) intrusions, (b) a correct word was recalled twice, and (c) participants did not recall three words at all. The last point was primarily relevant in the 1-week recall.
N = Negative words. P = Positive words. O = Neutral words.

(see Table F13 in the Appendix for an overview of all effects).[22]

For all participants, the number of correctly recalled words decreased from the learning phase to the 1-week recall. Thus, with longer retention intervals, participants generated less often three correct words within the first three words recalled. This effect was due to three reasons: With increasing retention interval, participants (a) generated more and more intrusions, (b) recalled a word more often twice, and (c) didn’t recall three words at all. However, older adults generated less often than younger adults correct responses from the first three words recalled. This age difference magnified with increasing retention interval.

There was no significant main effect of valence and no other significant interaction effect with valence. Follow-up comparisons did not reveal as well significant differences between valence categories (see Table F14 in the Appendix). Thus, there was no evidence for differential prioritization of positive, negative, or neutral material in the output order of the first three words recalled.

5.2.4.2 Summary

The analyses on the first three words recalled were instructed by the idea that the proposed positivity effect might be visible in the output order of the recalled words rather than in memory performance. That is, positive words might be recalled prior to negative and neutral words. The analyses did not support this idea.

[22] Table F13 in the Appendix provides as well the effects for an analysis on the first five words recalled. The analysis on the first five words recalled revealed similar effects as the analysis on the first three words. Thus, findings were not due to a specific output interval.

There was no evidence for a general prioritization of positive, negative, or neutral material in the sequence of recalling the words. Moreover, there was also no evidence for age-related differences in the recall of positive, negative, and neutral words within the first three words recalled. Thus, the findings did not support the idea that older adults showed a prioritization of positive words in the sequence of recalling emotionally-toned words.

5.2.5 Follow-up Analyses on Subjective Valence Categorization

Despite the attempt of the preparatory Word Rating Study in selecting an appropriate set of positive, negative, and neutral to-be-remembered words for the experiment, it might well be that the participant's evaluation of these words did not match to the a priori valence categorization. For example, a participant might evaluate a word as negative although this word was categorized as neutral based on the Word Rating Study. To examine the influence of participants' subjective evaluations of the to-be-remembered words, subjective valence categorizations were used. Thus, the recall performance in the emotion-heterogeneous list condition was reanalyzed with respect of each participant's self-generated valence categories.[23]

After the 1-week recall, participants were asked to rate all 90 words on a seven-point scale from *very unpleasant* (1) to *very pleasant* (7). Words scored as 1 or 2 were categorized as subjectively negative, words scored 3, 4, or 5 were categorized as subjectively neutral, and words scored 6 or 7 were categorized as subjectively positive for this participant. Table 26 provides the agreement in percent between participants' valence categorization and the a priori valence categorization for young and older adults. There was general agreement between subjective and a priori classification for young ($\kappa_W = .62$) and older adults ($\kappa_W = .69$). However, the agreement was somewhat larger for older than for young adults.[24] There were as well substantial cross-classifications signifying that the subjective evaluation of the to-be-remembered words differ for some participants from the a priori classification. This finding as important in two respects: First, a somewhat different pattern in the subjective than in a prior valence classifications was the prerequisite for a potentially different recall pattern for the subjective valence categorization. If there were 100% agreement between subjective and

[23] In the emotion-homogeneous list condition, this procedure was not appropriate due to restrictions in variability. In an emotion-homogeneous list, nearly all words should be related to one category.
[24] There was no evidence that any other person characteristic (i.e., education, crystallized and fluid intelligence) influenced the agreement between subjective and a priori classifications.

Table 26

Agreement between the A Priori Valence Categorization and the Subjective Valence Categorization in the Emotion-Heterogeneous List Condition

	Subjective Valence Categorization					
	Young Adults			Older Adults		
A Priori Classification:	Negative	Neutral	Positive	Negative	Neutral	Positive
Mean Percentages						
Negative	64.1	34.3	1.7	73.3	25.2	1.5
Neutral	16.5	68.1	15.4	21.5	63.0	15.5
Positive	0.8	25.6	73.5	0.3	16.7	83.0
Standard Deviations						
Negative	22.0	21.9	5.6	30.0	21.0	4.6
Neutral	14.3	20.6	13.7	21.1	32.1	12.9
Positive	2.8	23.8	23.6	1.8	23.5	24.6

a priori classifications, no differences in the memory pattern would be expected. Second, the cross-classifications indicated the importance of assessing the subjective evaluation of the to-be-remembered material in determining the impact of participants' evaluations.

The recall data was analyzed by computing the percent of recalled words within each participant' own subjective valence category based on the total number of words rated as either subjectively negative, positive or neutral. This was done for each trial in the learning and retention phase of the emotion-heterogeneous list condition. Table 27 provides mean percentages and their standard deviations. For example, the first value in the first row and column (i.e., 31.42%) indicates that young adults recalled on average 31.42% of the subjectively perceived negative words in Trial 1.

5.2.5.1 Subjective Valence Categories in the Learning Phase

The recall data based on subjectively generated valence categories were subjected to a 2 x 5 x 3 (Age x Trial x Valence) mixed analysis of variance with age (young vs. old) as a between-subjects factor and trial (5 learning trials) and subjective valence category (negative vs. positive vs. neutral) as two within-subjects factors. The analysis revealed significant main effects of age, $F(1,67) = 19.85$, $p < .001$, $\eta_p^2 = .23$; trial, $F(4,64) = 89.10$, $p < .001$, $\eta_p^2 = .85$; and subjective valence category, $F(2,66) = 10.20$, $p < .001$, $\eta_p^2 = .24$; and a significant interaction between age and trial, $F(4,64) = 3.85$, $p = .007$, $\eta_p^2 = .19$. There was no other significant effect, especially no significant interactions between age and subjective valence, $F(2,66) = 0.01$, $p = .997$, $\eta_p^2 < .01$; trial and subjective valence, $F(8,60) = 1.17$, $p = .331$, η_p^2

Table 27

Percent of Recalled Words within Subjectively Generated Valence Categories in the Heterogeneous List Condition for Young and Older Adults

	Mean Percentages						Standard Deviations					
	Young Adults			Old Adults			Young Adults			Old Adults		
	N	P	O	N	P	O	N	P	O	N	P	O
Learning Phase												
Trial 1	31.42	28.76	27.95	24.15	25.22	22.08	14.73	14.18	12.56	14.24	15.57	14.23
Trial 2	48.54	42.54	38.25	38.50	29.00	31.81	23.63	17.67	15.85	15.46	11.32	18.42
Trial 3	55.09	48.70	47.85	45.49	41.25	35.54	23.10	21.52	17.03	16.48	20.02	1821
Trial 4	63.03	41.47	37.66	50.44	41.47	37.66	21.40	18.59	23.14	17.83	21.22	23.14
Trial 5	67.79	63.08	61.63	50.64	47.39	45.66	22.21	18.58	20.93	18.28	19.88	21.56
1-Hour Recall												
Trial 1	56.53	49.38	49.17	36.43	31.41	30.37	26.55	18.82	22.29	21.54	18.99	23.29
Trial 2	55.58	47.87	47.46	37.95	31.81	28.82	23.35	21.78	21.84	23.91	18.78	21.13
Trial 3	57.99	51.53	47.87	36.16	32.73	28.86	24.71	22.70	23.34	22.04	17.95	23.15
1-Week Recall												
Trial 1	41.74	34.81	33.36	19.90	18.11	19.37	26.51	21.52	21.08	19.78	16.58	19.00
Trial 2	43.76	38.52	34.93	20.32	20.11	20.73	27.55	21.63	21.74	21.15	19.84	19.63
Trial 3	45.35	39.00	36.51	21.97	19.64	22.11	28.36	21.98	22.59	20.83	18.64	19.85

Note. N = Negative words. P = Positive words. O = Neutral words.

= .13; and between age, trial, and subjective valence, $F(8,60) = 0.69$, $p = .600$, $\eta_p^2 < .04$ (see Table F15 in the Appendix).

As expected, young adults recalled more words than older adults. All participants showed substantial increases in recall performance across the learning phase; however, young adults showed even steeper learning curves than older adults. This pattern of findings resembled the pattern of findings for the main analyses in the learning phase (see *5.2.1 Recall Performance in the Learning Phase*).

Follow-up comparisons between subjectively generated valence categories revealed that all participants learned more negative than neutral words, $F(1,67) = 20.51$, $p < .001$, $\eta_p^2 = .23$, and more negative than positive words, $F(1,67) = 7.03$, $p = .010$, $\eta_p^2 = .10$. The comparison between subjectively positive and subjectively neutral words was not significant, $F(1,67) = 1.81$, $p = .183$, $\eta_p^2 = .03$ (see Table F16 in the Appendix for details). The recall pattern for the subjective valence classification resembled the pattern for the a priori categorization (see section *5.2.1.1 Heterogeneous List Condition*). Moreover, the significant effect sizes were even stronger for the subjective (.23 and .10) than for the a priori valence categorization (.12 and .03). This might indicate that some amount of error variance involved in the analyses on the a priori valence categorization was due to interindividual differences in the subjective classification of positive, negative, and neutral words.

There was no evidence for age-related differences in remembering positive, negative, and neutral words in using subjectively generated valence categories. This was the case for the overall analysis as well as for the follow-up comparisons. (All effects were smaller than $\eta_p^2 < .001$.) Consistent with the main analyses on a priori valence categories, the analyses on subjective valence categories did not reveal any empirical support for the proposed positivity effect in older adults' memory.

5.2.5.2 Subjective Valence Categories in the Retention Phase

Similar to the main analyses for the retention phase (see section *5.2.2 Recall Performance in the Retention Phase*), I conducted a 2 x 3 x 3 (Age x Time x Valence) mixed analysis of variance with age (young vs. old) and subjective valence category (negative vs. positive vs. neutral) as between-subjects factors and time (learning trial 5 vs. 1-hour vs. 1-week) as a within-subjects factor. The analyses yielded significant main effects of age, $F(1,67) = 22.08$, $p < .001$, $\eta_p^2 = .25$; time, $F(2,66) = 137.76$, $p < .001$, $\eta_p^2 = .81$; and subjective valence, $F(2,66) = 4.99$, $p = .010$, $\eta_p^2 = .13$. No other effect reached the significance level. In particular, there were no significant interactions of Age x Valence, $F(2,66) = 0.46$, $p = .663$, $\eta_p^2 = .01$; Age x Trial, $F(2,66) = 0.60$, $p = .552$, $\eta_p^2 = .02$; Trial x Valence, $F(4,64) = 0.74$, $p = .572$, $\eta_p^2 = .04$; and Age x Trial x Valence, $F(4,64) = 0.69$, $p = .600$, $\eta_p^2 = .04$ (see Table F15 in the Appendix).

As expected, young and older adults showed a considerable amount of forgetting across the 1-week retention period. However, older adults recalled fewer words than young adults. The follow-up comparisons between subjective valence categories confirmed the pattern in the learning phase: Participants recalled more negative than neutral words, $F(1,67) = 9.63$, $p = .013$ $\eta_p^2 = .13$, and more negative than positive words, $F(1,67) = 4.56$, $p = .036$, $\eta_p^2 = .06$. The comparison between subjectively positive and subjectively neutral words was not significant, $F(1,67) = 0.44$, $p = .508$, $\eta_p^2 = .01$ (see Table F16 in the Appendix for details). For both age groups, the right part of Figure 8 depicts percentages for subjective positive, negative, and neutral words in the learning phase, the 1-hour recall, and the 1-week recall. To facilitate comparisons to the main analyses, the percentages from a priori valence categories are depicted in the left part of Figure 8.

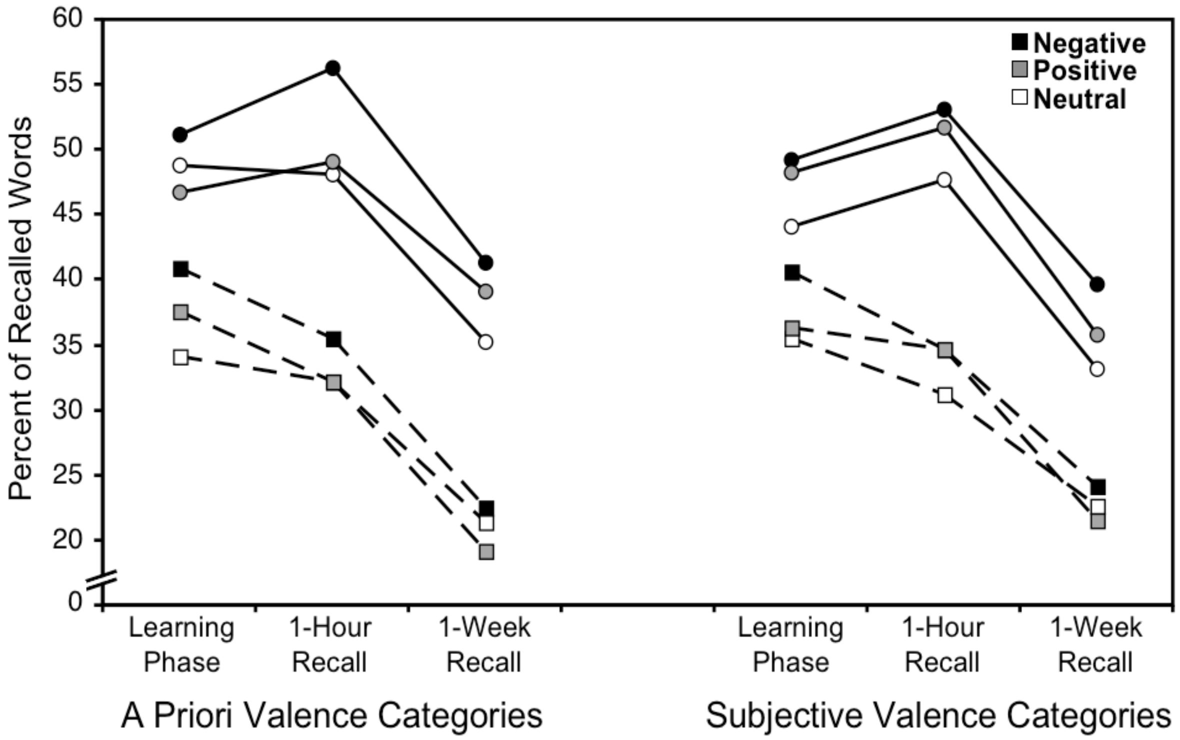

Figure 8 Mean recall percentages for a priori valence categories (left part) and subjective valence categories (right part) as a function of age group and retention interval.

For the retention phase, the pattern of findings for subjectively generated valence categories resembled the pattern of findings for a priori valence categories. However, the effect sizes for recalling more negative than positive and neutral words (i.e., negativity effect) were somewhat larger for the analyses on subjective valence categories (.13 and .06) than on a priori valence categories (.11 and .05). Regarding age, there was no evidence for age-related differences in long-term memory for subjective positive, negative, and neutral words. Specifically, there was no evidence for the proposed positivity effect in older adults' long-term memory using subjectively generated valence categories.

5.2.5.3 Summary

Consistent with Hypothesis 6, the analyses on subjectively generated valence categories supported the findings from the main analyses on a priori valence categories in the emotion-heterogeneous list condition. Young and older adults recalled more negative than positive words and more negative than neutral words in the learning and retention phase. Despite some differences between the subjective and a priori valence classification, the recall pattern was consistently found for the subjective and a priori valence classification.

Moreover, the effect sizes were even somewhat larger for the subjective than for the a priori valence categorization. This might indicate that some amount of error variance in the analyses on a priori valence categories was due to interindividual differences in the perceived valence of the to-be-remembered words.

Consistent with the main analyses on a priori valence categories, the analyses on subjective valence categories provided no evidence for age-related differences in the positive-negative disparity. In particular, there was no evidence for an advantage of older adults' memory for positively-toned words.

5.2.6 Analyses on Word Sets Across Conditions

The separate analyses in the heterogeneous and homogeneous list conditions suggested that the pattern of valence effects differ by condition indicating the proposed heterogeneity-homogeneity distinction. The different data structure in both conditions (valence was a between-subject factor in the homogeneous list condition and a within-subject factor in the heterogeneous list condition) did not allow a direct comparison between conditions. However, in both conditions, identical sets of 10 words were used and these were balanced across trials and participants in parallel. Therefore, it was possible to compare the recall of the 10 items in the same word set in the different list contexts (i.e., recall for the same 10 negative words in the context of other negative words in the emotion-homogeneous list condition versus in the context of positive and neutral words in an emotion-heterogeneous list condition). For each 10-word set, I calculated a recall discrepancy score by subtracting the mean item recall in a heterogeneous list condition from mean item recall in a homogeneous list condition separately for young and older adults. This was done for each trial separately for young and old adults.

The analysis of word set recall discrepancy scores yielded a direct test of whether there was differential processing in the heterogeneous and homogeneous conditions. If differential processing prioritization contributed to remembering emotional material either positive or negative, word set recall discrepancy scores should be significantly different from zero and vary across the valence categories. The following sections are divided in the analysis for the learning phase and the analysis across the retention phase (including the learning phase).

5.2.6.1 Word Sets in the Learning Phase

For both age groups, recall discrepancy scores were calculated for all word sets (3 per valence category) and all five learning trials. This procedure resulted in 15 (3 sets x 5 trials) values for each valence category and age group (see Table F17 in the Appendix for means and standard deviations). Word set recall discrepancy scores were subjected to a 3 x 2 (Valence x Age) mixed ANOVA with valence as a between-sets factor and age as a within-sets factor. Thus, the units of analysis were word sets rather than participants. The analysis revealed significant main effects for age, $F(1,42) = 20.36$, $p < .001$, $\eta_p^2 = .33$, and valence, $F(2,42) = 16.32$, $p < .001$, $\eta_p^2 = .44$, and a significant interaction between age and valence, $F(2,42) = 16.31$, $p < .001$, $\eta_p^2 = .44$.[25]

As Figure 9 indicates, compared with the emotion-homogeneous list condition, young adults recalled from the set of 10 negative words one additional word and from the set of 10 neutral words one additional word in the emotion-heterogeneous list condition on average. In the emotion-homogeneous list condition, however, young adults recalled from the set of 10 positive words on average 0.5 more words than in the emotion-heterogeneous list condition (all contrasts were significant at $p < .01$). For older adults, the general recall discrepancy pattern was similar to the pattern in young adults. However, recall discrepancy scores for the

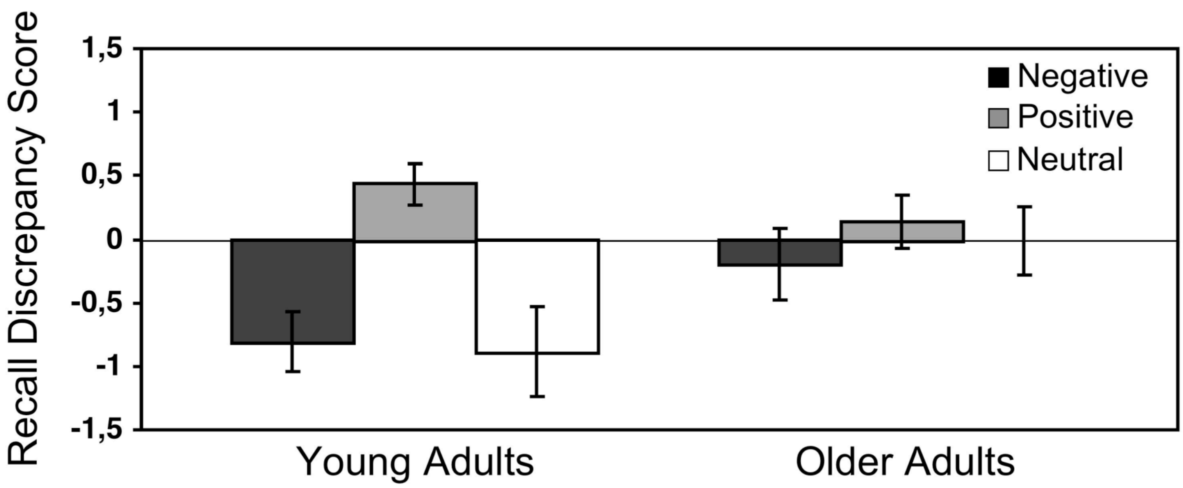

Figure 9 Recall discrepancy scores for young and older adults for the same 10-item sets of negative, positive, and neutral words presented in the homogeneous and heterogeneous list contexts. Values greater than zero indicate that words were better recalled in an emotion-homogeneous list condition, whereas values less than zero indicate that words were better recalled in an emotion-heterogeneous list condition. Error bars represent 95% confidence intervals.

[25] Including trial (five learning trials) as an additional between-sets factor in the mixed analysis of variance did not reveal any effect of trial.

valence categories for older adults were not significantly different from zero indicating equal recall for negative, positive, and neutral words across conditions (all contrasts were nonsignificant, $p > .20$). This is evidence that older adults' memory for emotionally-toned material was less affected by differential processing prioritization.

5.2.6.2 Word Sets in the Retention Phase

Recall discrepancy scores were calculated for each trial in the retention phase. This procedure yielded 9 (3 sets x 3 trials) values for each retention interval, valence category, and age group. These scores were arranged with the recall discrepancy scores from the learning phase into one overall analysis. A 2 x 3 x 3 (Age x Time x Valence) mixed analysis of variance with valence (negative vs. positive vs. neutral) as a between-sets factor and age (young vs. old) and time (learning vs. 1-hour vs. 1-week) as within-sets factors revealed significant main effects for age, $F(1,90) = 63.58$, $p < .001$, $\eta_p^2 = .41$, and valence, $F(2,90) = 62.79$, $p < .001$, $\eta_p^2 = .58$, and significant interactions for Age x Valence, $F(2,90) = 22.92$, $p < .001$, $\eta_p^2 = .34$; Age x Time, $F(2,90) = 6.86$, $p = .002$, $\eta_p^2 = .13$; and Age x Time x Valence, $F(4,90) = 2.50$, $p = .048$, $\eta_p^2 = .10$. The main effect of time, $F(2,90) = 2.97$, $p = .056$, $\eta_p^2 = .06$, and the interaction between time and valence, $F(4,90) = 2.06$, $p = .093$, $\eta_p^2 = .08$, were - marginally but - non-significant.

For the young adults, the recall discrepancy pattern was quite consistent across the learning phase and the retention intervals. Young adults recalled more negative and more neutral words in the heterogeneous list condition than in the homogeneous list condition (~ 3/4 to 1 word); but they recalled more positive words in the homogeneous list condition than in the heterogeneous list condition (~ 1/2 to 1 word). For the older adults, the recall discrepancy pattern changed across the retention phase. As mentioned in the previous section, recall discrepancy scores for older adults were not significantly different from zero in the learning phase. However, in the 1-hour and the 1-week recall, older adults recalled more positive words in the homogeneous than in the heterogeneous list condition (~ 3/4 to 1 word). Moreover, in the 1-week recall, older adults recalled also more negative words in the homogeneous than in the heterogeneous list condition (~ 1/2 word). Separately for young and older adults, Figure 10 shows recall discrepancy scores for the learning phase, the 1-hour recall, and the 1-week recall.

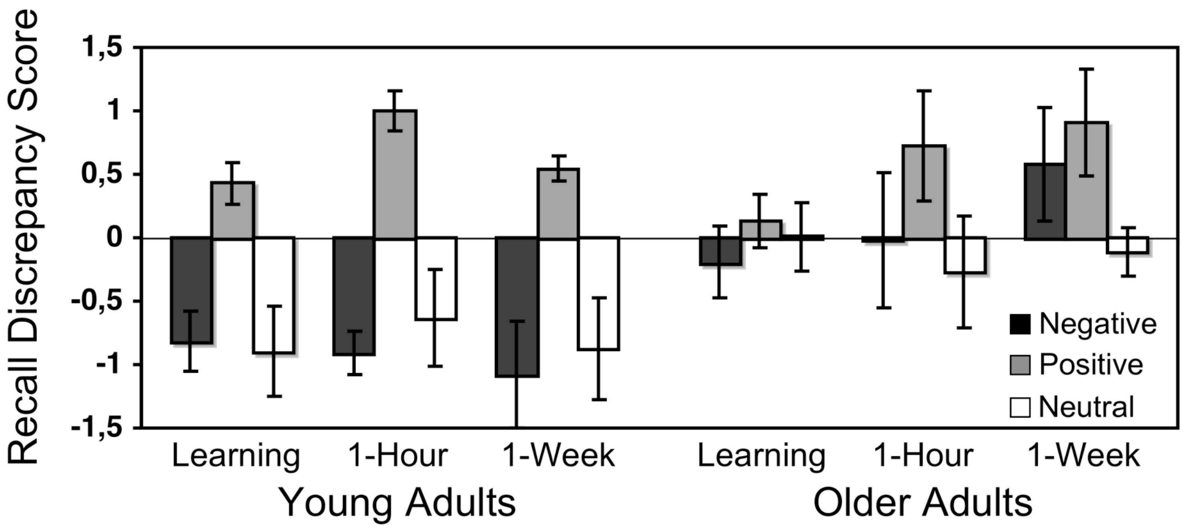

Figure 10 Recall discrepancy scores for young and older adults in the learning phase and in both retention intervals. Values greater than zero indicate that word sets were better recalled in an emotion-homogeneous list condition, whereas values less than zero indicate that word sets were better recalled in an emotion-heterogeneous list condition. Error bars represent 95% confidence intervals.

5.2.6.3 Word Sets: Summary

The different recall patterns found in the separate analyses for emotion-heterogeneous and emotion-homogeneous list conditions suggested a heterogeneity-homogeneity list distinction (see section *5.2.1* and *5.2.2*). The incompatible data structure in both conditions did not permit a direct comparison of both conditions. The analysis of recall of 10-item word sets, however, provided the opportunity for a direct test between conditions.

In the learning phase, young adults showed strong contextual effects: They recalled more negative and neutral words and fewer positive words in the heterogeneous list condition than in the homogeneous list condition. The overall pattern was similar for older adults but, unlike the young adults, their recall discrepancy scores were not significantly different from zero. This suggests that older adults did not use emotion-based processing prioritization as much as young adults.

In the retention phase, the recall discrepancy pattern on 10-item word sets was very consistent for young adults. Young adults recalled more negative and neutral words in a heterogeneous list conditions but more positive words in a homogeneous list condition. The recall discrepancy pattern for older adults changed across the retention period. Beginning with no evidence for differential recall in the heterogeneity-homogeneity comparison in the learning phase, older adults recalled in the homogeneous list condition more positive and more negative words than in the heterogeneous list condition in the 1-week retention interval.

Together with the means of recall performance, this indicated that older adults could maintain a higher number of words in an emotion-homogeneous list context with only negatively-toned or only positively-toned words than in a corresponding emotion-heterogeneous list context.

5.2.7 Follow-up Analyses on Recallability

The analyses in the previous section (see section *5.2.6*) were done on the level of 10-item word sets. In this section, the analyses focused on the level of single words and their corresponding characteristics. To examine the association between memory performance and word characteristics, word *recallability* scores were computed. Recallability is defined as the number of times a word was actually recalled across participants and trials (Rubin & Friendly, 1986). The analyses were instructed by the research questions of whether word characteristics (other than valence) had an impact on remembering positive, negative, and neutral words.

The analyses on the level of single words and the corresponding recallability scores had two advantages: First, similar to the analyses of 10-item word sets, these analyses permitted to analyze both list conditions concurrently. Second, and more important, the analyses on the level of single words allowed entering word characteristics as covariates. Before describing the analyses on recallability scores, I provide some general information about the word ratings assessed in the experiment especially their correlations to ratings from the word rating study.

5.2.7.1 Inter-Correlations for Ratings from the Word Rating Study and the Experiment

To check whether word ratings from the experiment were consistent with word ratings from the Word Rating Study (see Chapter 4), correlations between word ratings were examined. In the experiment, word ratings were assessed on all 90 to-be-remembered words. The rating dimensions were valence, arousal, young-stereotype, old-stereotype, and self-relevance. Means and standard deviations for the rating data obtained in the experiment are provided in Table F19 in the Appendix. Analyses of variance are listed in Table F20 in the Appendix. For each word, means and standard deviations are presented in Appendix G including overall means for each dimension (see Table G1) as well as separate means for both age groups and list conditions with corresponding analyses of variance (see Table G2-G6).

Similar to the Word Rating Study, a derived score of emotional intensity was computed as the absolute deviation from the neutral midpoint of the valence dimension. For all word ratings from the experiment, Table 28 provides inter-correlations as well as correlations with ratings from the Word Rating Study including the objective measures of word frequency, word frequency class, and word length.

The pattern of inter-correlations within the experiment was similar to the pattern in the Word Rating Study. Ratings of valence were highly correlated to ratings of arousal (*r* = -.82) and self-relevance (*r* = .93). Negative words were more arousing than positive words whereas positive words were more typically for one's self than negative words. Moreover, valence was also associated to ratings of young-stereotype (*r* = .32) and old-stereotype (*r* = .45). Words that were either typical for young adults or typical for older adults were more positive. Arousal was also related to ratings of old-stereotype (r = -.48) indicating that attributes of older adults were perceived as low-arousing.

Table 28

Correlations between Ratings from the Word Ratings Study and from the Experiment

	Experimental Sample					
	V	EI	A	Y-S	O-S	S
Objective Measures						
Word Frequency	.18	.00	-.10	.02	.14	.16
Word Frequency Class	-.14	.04	.11	.06	-.13	-.11
Word Length	.06	-.06	.09	.24*	.10	.14
Experimental Sample						
Valence (V)	1.00	.11	-.82***	.32**	.45***	.93***
Emotional Intensity (EI)	.11	1.00	-.03	.14	-.00	.01
Arousal (A)	-.82***	-.03	1.00	.03	-.48***	-.72***
Young-Stereotype (Y-S)	.32**	.14	.03	1.00	-.12	.41***
Old-Stereotype (O-S)	.45***	-.00	-.48***	-.12	1.00	.55***
Self-Relevance (S)	.93***	.01	-.72***	.41***	.55***	1.00
Word Rating Study Sample						
Valence	**.99*****	.09	-.82***	.32**	.44***	.92***
Emotional Intensity	.13	**.93*****	-.02	.14	.06	.07
Arousal	-.68***	-.01	**.94*****	.06	-.41***	-.58***
Control	.61***	.12	-.29*	.35***	.31**	.64***
Imagery	.09	.27*	.02	.33**	-.15	-.02
Age-Relevance	-.03	-.11	-.24*	**-.71*****	**.69*****	-.03
Self-Relevance	.91***	.04	-.68***	.44***	.54***	**.97*****

Note. Values in bold documented retest-reliabilities and should be high.
* $p < .05$. ** $p < .01$. *** $p < .001$.

The correlations between the experiment and the word rating study revealed high consistencies in the rank ordering of the words (all over .90). Especially the correlation between ratings of valence was very high between studies (r = .99). These high correlations gave support for the reliability of the word ratings; in particular, it confirmed the reliability of the valence ratings and their validity as basis for word categorization. Again, word ratings were generally not related to objective measures of word frequency, word frequency class, and word length. One exception was a small correlation between word length and young-stereotype in the experimental sample (r = .24) indicating that typical attributes of young adults were somewhat longer than non-typical words.

5.2.7.2 Recallability Scores and Their Associations to Word Characteristics

Recallability scores were calculated by counting the number of times a word was recalled across trials and participants. For example, overall recallability scores across the learning phase were made by counting the number of times a word was recalled in all five learning trials from all participants who had to learn this word. This procedure for the learning phase resulted in a maximum score of 240 (5 trials x 2 conditions x 2 age groups x 12 participants). Tables G7-G9 in the Appendix provides recallability scores for the 30 negative, 30 neutral, and 30 positive words respectively. These tables include separate scores for young and older adults as well as for heterogeneous and homogeneous list conditions.

Recallability scores were linked to three sources of word characteristics: (a) the three objective measures of word frequency, word frequency class, and word length; (b) the rating data obtained in the Word Rating Study; and (c) the rating data obtained in the experiment. Table 29 presents the correlations separately for both age groups, for heterogeneous and homogeneous list conditions, and for the learning phase, the 1-hour and the 1-week recall.

The correlation pattern can be summarized into four main findings: First, independent from age group, list condition, and retention interval, word length was systematically related to recallability scores ($-.27 \leq r \leq -.50$). Shorter words were more often recalled than longer words. Second, recallability scores showed positive correlations with imagery ($.08 \leq r \leq .23$) and emotional intensity ($.03 \leq r \leq .31$). Easy to imagine words were more often recalled than hardly to imagine words. Emotional words were more often recalled than neutral words. However, correlations were generally small and did not exceed the significance level in many cases. Third, for valence and arousal, the magnitude and the direction of the correlations varied as a function of list condition and age group. Recallability was only significantly correlated with valence ($.20 \leq r \leq .26$) and arousal ($-.17 \leq r \leq -.25$) for young adults in the

Table 29

Correlations between Recallability Scores and Other Word Characteristics

	Young Adults						Older Adults					
	Heterogeneous			Homogeneous			Heterogeneous			Homogeneous		
	L	H	W	L	H	W	L	H	W	L	H	W
Objective Measures												
Frequency	-.16	-.11	-.10	-.03	.02	.05	-.01	.06	.16	-.04	.03	.00
Frequency Class	.05	.04	.03	-.06	-.07	-.11	-.07	-.07	**-.22**	-.03	-.08	-.04
Word Length	**-.39**	**-.38**	**-.43**	**-.35**	**-.29**	**-.27**	**-.45**	**-.50**	**-.40**	**-.38**	**-.34**	**-.31**
Experimental Sample												
Valence	-.12	-.14	-.06	.20	**.22**	**.26**	-.13	-.13	-.13	-.04	.04	-.04
Emotional Intensity	.11	.18	.20	**.26**	**.28**	**.30**	**.32**	.19	.08	**.24**	**.28**	**.28**
Arousal	.05	.09	.09	**-.25**	**-.22**	**-.21**	.07	.07	.11	-.06	-.12	.02
Young-Stereotype	-.10	-.07	.00	-.01	.00	.00	-.08	-.08	-.04	-.08	-.04	.01
Old-Stereotype	**-.22**	-.12	-.19	-.01	.04	.04	-.17	-.13	-.16	-.03	.02	-.03
Self-Relevance	-.16	-.16	-.10	.12	.14	.19	**-.22**	-.20	-.17	-.11	-.04	-.09
Word Rating Study Sample												
Valence	-.10	-.12	-.04	**.22**	**.23**	**.26**	-.12	-.12	-.11	-.02	.06	-.02
Emotional Intensity	.10	**.21**	**.22**	**.26**	**.27**	**.31**	**.27**	.15	.03	**.22**	**.26**	**.30**
Arousal	.04	.07	.13	**-.22**	-.19	-.17	.03	.04	.10	-.09	-.14	.02
Control	.03	.04	.12	.14	**.25**	**.35**	-.05	.04	.07	.00	.04	.15
Imagery	.12	**.22**	.18	.12	.15	.14	**.23**	**.23**	.16	.11	.16	.08
Age-Relevance	-.10	-.05	-.12	-.04	-.04	-.04	-.02	.02	-.04	.05	.04	-.03
Self-Relevance	-.15	-.16	-.09	.14	.16	.20	-.19	-.17	-.15	-.07	-.01	-.06

Note. Correlations in bold were significant at $p < .05$. L = Learning Phase. H = 1-Hour Recall. W = 1-Week Recall.

emotion-homogeneous list condition: Positive words and low-arousing words were more often recalled than negative words and high-arousing words. For the emotion-heterogeneous list condition, corresponding correlations with recallability changed the direction of association for valence ($-.04 \leq r \leq -.14$) and arousal ($.03 \leq r \leq .13$). Again, these correlations were small and did not exceed the significance level. Finally, word frequency, word frequency class, control, age-relevance, young-stereotype, old-stereotype, and self-relevance were not systematically associated to recallability scores. Self-relevance seemed to resemble the pattern for valence but did not generally reach the significance level.

Taken together, subjective ratings showed only small and most often non-significant correlations with recallability scores. The only dimension that showed relative consistent and high correlations with recallability was emotional intensity, the derived score based on valence ratings. Word length as one of the objective measures showed considerably large correlations with recallability.

5.2.7.3 Differences between Age Groups and List Conditions in Recallability

Recallability scores derived from all 90 to-be-remembered words in the learning phase were dedicated to a mixed analysis of variance with valence category (negative vs. neutral vs. positive) as a between-words factor and time (learning phase vs. 1-hour recall vs. 1-week recall), age (young vs. older adults) and list condition (heterogeneous vs. homogeneous list condition) as within-words factors. The analysis revealed significant main effects of age, $F(1,87) = 90.94$, $p < .001$, $\eta_p^2 = .511$, and list condition, $F(1,87) = 5.18$, $p = .025$, $\eta_p^2 = .056$, as well as significant interactions between age and list condition, $F(1,87) = 7.41$, $p = .008$, $\eta_p^2 = .079$; valence and list condition, $F(2,87) = 6.77$, $p = .002$, $\eta_p^2 = .135$; and valence, age, and list condition, $F(2,87) = 5.94$, $p = .004$, $\eta_p^2 = .120$. The main effect of valence, $F(2,87) = 0.97$, $p = .383$, $\eta_p^2 = .022$, and the interaction between age and valence, $F(2,87) = 1.28$, $p = .283$, $\eta_p^2 = .029$, did not reach the significance level (see Table F21 in the Appendix for an overview of all effects). For both age groups and list conditions, Table 30 provides means and standard deviations of recallability for positive, negative, and neutral words in the learning phase.

Generally, words were recalled more often by young than by older adults. Moreover, words were more often recalled in the heterogeneous list condition than in the homogeneous list condition by young adults whereas words were equally well recalled by older adults in the heterogeneous and homogeneous list condition.

The valence interactions were examined with follow-up comparisons between valence categories. For positive and neutral words, these analyses revealed significant interaction effects of Valence x Condition, $F(1,58) = 9.02$, $p = .004$, $\eta_p^2 = .135$, and of Valence x Age x Condition, $F(1,58) = 9.02$, $p < .001$, $\eta_p^2 = .191$, signifying that positive words were recalled more often in an emotion-homogeneous list context than in an emotion-heterogeneous list context. Moreover, positive words were recalled more often and neutral words less often in

Table 30

Recallability in the Learning Phase by List Condition, Age Group, and Valence Category

	Means				Standard Deviations			
	Young Adults		Older Adults		Young Adults		Older Adults	
	M	Pure	Mixed	Pure	Mixed	Pure	Mixed	Pure
Negative Words	30.67	25.80	24.53	23.37	8.88	7.51	8.08	8.32
Positive Words	28.03	30.93	22.53	23.33	12.06	9.57	8.17	9.31
Neutral Words	29.27	23.93	20.47	20.47	10.27	10.59	9.32	9.79

Note. M = Mixed, Heterogeneous List Condition. P = Pure, Homogeneous List Condition.

an emotion-homogeneous than in an emotion-heterogeneous list context by young adults, whereas positive and neutral words were similarly well recalled across both conditions by older adults. The follow-up comparison between positive and negative words revealed as well significant interaction effects for Valence x Condition, $F(1,58) = 10.86$, $p = .002$, $\eta_p^2 = .158$, and for Valence x Age x Condition, $F(1,58) = 5.05$, $p = .028$, $\eta_p^2 = .080$, demonstrating that positive words were recalled more often and negative words less often in an emotion-homogeneous list context by young adults whereas positive and negative words were equally well recalled in both list conditions by older adults. The follow-up comparison between negative and neutral words did not reveal any significant interaction with valence (see Table F22 in the Appendix for details).

In sum, the pattern of findings for the recallability scores resembled the pattern of findings from the main analyses in the learning phase (see section *5.2.1 Recall Performance in the Learning Phase*) and the analyses on 10-item word sets (see section *5.2.6 Analyses on Word Sets across Conditions*). The analyses revealed direct evidence that valence category per se was not significantly associated to memory performance (i.e., no main effect of valence) but rather list context modulated the effect of valence on memory performance (i.e., List x Valence). Thus, processing priority for negative words in emotion-heterogeneous but not in emotion-homogeneous list contexts. Consistent with the analyses on 10-item word sets, the analyses on single words revealed an Age x List x Valence interaction signifying that older adults showed less differential processing prioritization for emotionally-toned words in both list contexts.

In contrast to the previous analyses, the analyses on word recallability were done on the single word level and not on the person level. This allowed me to analyze the influence of word characteristics on remembering emotionally-toned words.

5.2.7.4 Word Characteristics as Covariates

The main goal of the analyses on word recallability was to investigate the influence of word characteristics (other than valence) on remembering positive, negative, and neutral words in the learning phase. To do this, analyses of covariance were performed with valence category (negative vs. neutral vs. positive) as a between-words factor and age (young vs. older adults) and list condition (heterogeneous vs. homogeneous list condition) as within-words factors.

Word covariates were obtained from three sources: (a) the objective measures of word characteristics (i.e., word frequency, word frequency class, word length); (b) the rating data obtained in the Word Rating Study (i.e., valence, emotional intensity, arousal, control, imagery, age-relevance, self-relevance); and (c) the rating data obtained in the experiment (i.e., valence, emotional intensity, arousal, young-stereotype, old-stereotype, self-relevance). In this context, the objective measures and the rating information from the Word Rating Study were constant covariates. The rating information from the experiment, however, can be linked to each level of the within-subjects factors and were varying covariates. As described above (see section *5.2.3.2 Person Characteristics as Covariates*), constant covariates cannot affect the impact of within-subjects effects but varying covariates can. Thus, constant word covariates cannot change the effects of both within-subjects factors: age group and list condition. However, constant covariates can alter the main effect of valence.

For all 16 word covariates, analyses of covariance were conducted. Table 31 provides effect sizes for the main effects of age (A), list condition (L), and valence category (V), as well as for the interactions of age and condition (AL), age and valence (AV), condition and valence (LV), and age, condition, and valence (ALV). Only effect sizes are depicted to facilitate the comparison between different analyses (and to reduce visual complexity of the table). The effect sizes for the analyses without a covariate are shown in the top of Table 31 (see section *5.2.7.3*). The changes in effect sizes for the analyses of covariates are informative for the relative impact of other word characteristics on the obtained effects (rather than the valence category alone).

The analyses of covariance did not provide any systematic evidence that other word characteristics (rather than valence category) had an impact on the obtained pattern of findings in the present dissertation. In particular, the interactions between list condition and valence as well as between age, list condition, and valence were still significant after adjustment for word characteristics. This document that age-related differences in differential processing prioritization for emotionally-toned words were not due to other word characteristics than valence. The non-significant main effect of valence and the non-significant interaction between age and valence did not reach the significance level by entering word covariates into the analysis. Thus, there was no evidence that neither valence per se nor an age by valence interaction was associated to recall performance. Again, the analyses of covariance on the single word level did not reveal a hidden positivity effect in older adults' memory.

Table 31

Effect Sizes by Entering Word Covariates in the Analyses of Word Recallability in the Learning Phase

Covariates	A	L	V	AL	AV	LV	ALV
Without Covariates	**.51**	**.06**	.02	**.08**	.03	**.14**	**.12**
Objective Measures							
Frequency	**.51**	**.06**	.02	**.08**	.03	**.14**	**.12**
Frequency Class	**.51**	**.06**	.02	**.08**	.03	**.14**	**.12**
Word Length	**.51**	**.06**	.02	**.08**	.03	**.14**	**.12**
Word Rating Study							
Valence	**.51**	**.06**	.02	**.08**	.03	**.14**	**.12**
Emotional Intensity	**.51**	**.06**	.02	**.08**	.03	**.14**	**.12**
Arousal	**.51**	**.06**	.03	**.08**	.03	**.14**	**.12**
Control	**.51**	**.06**	.02	**.08**	.03	**.14**	**.12**
Imagery	**.51**	**.06**	.01	**.08**	.03	**.14**	**.12**
Age-Relevance	**.51**	**.06**	.02	**.08**	.03	**.14**	**.12**
Self-Relevance	**.51**	**.06**	.05	**.08**	.03	**.14**	**.12**
Experiment							
Valence	**.51**	**.06**	.04	**.07**	.03	**.07**	**.12**
Emotional Intensity	**.50**	**.06**	.00	**.08**	.03	**.15**	**.12**
Arousal	**.51**	**.05**	.03	**.08**	.03	**.10**	**.12**
Young-Stereotype	**.51**	.04	.03	**.05**	.04	**.14**	**.12**
Old-Stereotype	**.48**	.03	.03	**.09**	.02	**.09**	**.09**
Self-Relevance	**.42**	**.06**	.06	**.07**	.03	**.15**	**.12**

Note. Effect sizes in bold were significant at $p < .05$. A = Age Group (young vs. old). L = List Condition (heterogeneous vs. homogeneous list). V = Valence (negative vs. positive vs. neutral).

5.2.7.5 Summary

The analyses on recallability scores supported the findings in the main analyses. Especially the interactions with valence resembled the findings on recall discrepancy scores by analyzing the recall of word sets (see section 5.2.3), that is, the heterogeneity-homogeneity distinction and the age differences found for recall discrepancy scores. Thus, for young adults, positive words were more often recalled in an emotion-homogeneous list condition and negative and neutral words were more often recalled in an emotion-heterogeneous list condition. In contrast, older adults showed a similar recall pattern in both conditions. In this respect, the analyses on the level of words revealed a similar pattern as the analyses on the level of persons.

The following analyses of covariance did not reveal any effect of word characteristics on the obtained findings. Specifically, the findings were practically unchanged in terms of effect sizes. This was not surprising given that the major findings relied on the distinction between the emotion-heterogeneous and emotion-homogeneous list condition. Word characteristics, however, applied to both conditions in the same way. Thus, word covariates could not explain context effects such as the heterogeneity-homogeneity distinction.