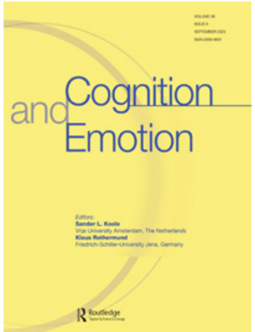

Cognition and Emotion

ISSN: 0269-9931 (Print) 1464-0600 (Online) Journal homepage: www.tandfonline.com/journals/pcem20

Emotion-enhanced source memory: effects of age and experimental setting

Nikoletta Symeonidou

To cite this article: Nikoletta Symeonidou (04 Aug 2025): Emotion-enhanced source memory: effects of age and experimental setting, Cognition and Emotion, DOI: 10.1080/02699931.2025.2541766

To link to this article: https://doi.org/10.1080/02699931.2025.2541766

Published online: 04 Aug 2025.

COGNITION AND EMOTION
https://doi.org/10.1080/02699931.2025.2541766

Routledge
Taylor & Francis Group

OPEN ACCESS

Check for updates

Emotion-enhanced source memory: effects of age and experimental setting

Nikoletta Symeonidou

Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim, Germany

ABSTRACT

This preregistered research examined whether younger and older adults robustly show enhanced source memory for socio-emotional versus neutral sources in a lab- and online-based experimental setting. The Lab Experiment ($N = 138$) was conducted in a lab room with German-speaking younger and older adults, while the Online Experiment ($N = 136$) was run on Prolific with English-speaking younger and older samples. In both experiments, neutral faces (= items) were shown on positive, negative, or neutral scenes (= sources) and participants rated at encoding how (un)pleased the face appeared. All stimuli were selected based on valence and arousal norms, ensuring their intended socio-emotional character. Memory was measured with a multinomial model. Across experiments, participants rated faces as least pleased when paired with negative scenes, moderately pleased with neutral scenes, and most pleased with positive scenes. Lab-recruited older adults additionally exhibited a positivity bias, which, however, was absent online. Source memory results revealed that younger adults in both experiments did not benefit from emotional sources. In contrast, lab-recruited older adults showed better source memory for emotional, especially positive, sources; this benefit was, however, absent online. Lower compliance and distractibility in the Online Experiment are discussed as explanations for the diverging results.

ARTICLE HISTORY
Received 24 July 2024
Revised 25 June 2025
Accepted 25 July 2025

KEYWORDS
Source memory; emotion-enhanced memory; aging; experimental setting; multinomial modelling

Aging comes along with worse episodic memory and particularly with worse source memory, that is, memory for the origin or context of information (e.g. temporal, spatial or social features; Johnson et al., 1993). For example, many studies have shown that older compared to younger adults have greater difficulty in remembering the speaker or spatial context of information while their memory for the central information per se (= item memory) is less affected (see Old & Naveh-Benjamin, 2008, for a meta-analysis). Although this age-related source memory deficit seems to rely on brain changes (Mitchell & Johnson, 2009), research has also shown that it is partially reducible. For example, older adults' source memory seems to improve with emotionally meaningful compared to neutral sources (May et al., 2005; Rahhal et al., 2002). However, results in this matter are inconsistent: Some studies show that the emotional benefit is more pronounced in older than in younger adults (May et al., 2005), while other studies (with better-controlled emotional material) show that older adults benefit less (Davidson et al., 2006) or not at all from emotional sources (Symeonidou et al., 2022). In two experiments, the present research reinvestigated age effects in emotional source memory by extending prior work through integrating key methodological strengths – specifically, by using normed emotional material and socially relevant item-source pairs. One experiment was conducted in the lab with a German-speaking sample.

CONTACT Nikoletta Symeonidou nsymeoni@mail.uni-mannheim.de Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim D-68161, Germany

Supplemental data for this article can be accessed online at https://doi.org/10.1080/02699931.2025.2541766.

Using the same material and procedure, another experiment was conducted in an online environment with English-speaking participants. This not only aimed to provide a valuable robustness check, but also to address persistent concerns about lower data quality in online experimental research (Chmielewski & Kucker, 2020; but see Uittenhove et al., 2023). Moreover, given that direct lab–online comparisons using identical paradigms are relatively rare in aging research, this paper offers a useful addition to the existing literature.

Aging and emotional memory

Several studies have shown that aging worsens episodic memory, presumably due to an age-related associative deficit (Naveh-Benjamin, 2000). That is, older adults have specific difficulties in binding different event features into a coherent unit. Age-related memory differences are thus larger in tasks measuring associative and source memory than item memory because performance in these tasks depends on successful binding of information (see Old & Naveh-Benjamin, 2008, for a meta-analysis). Consequently, research has explored ways to reduce this associative deficit, with promising results. For example, instructing older adults to focus on the item-source relationship (Glisky et al., 2001) or to use mediators that combine item and source (e.g. a sentence or image; Kuhlmann & Touron, 2017) considerably attenuates their source-memory deficit. Notably, using emotionally meaningful instead of neutral source material has also been shown to eliminate age effects in source memory (May et al., 2005). This latter finding is particularly intriguing because it does not require additional (experimenter-given) instructions. It also aligns with the socioemotional selectivity theory (SST; Carstensen, 2006), an influential life-span theory on emotion and motivation. The SST suggests that, as we age, our priorities shift from knowledge-oriented goals to emotion-oriented goals. More specifically, the theory assumes that the shortened time horizon in older age leads to an increased awareness of our mortality. This, in turn, makes emotionally meaningful pursuits more salient and important. Accordingly, the SST predicts that older compared to younger adults prioritise and therefore better remember emotionally meaningful and especially positive information (so-called age-related positivity effect, see Reed et al., 2014, for a meta-analysis). Most research on these emotion-based benefits in older adults' memory has exclusively focused on item memory, which is surprising given that their main deficit concerns associative memory, including source memory. May and colleagues (2005) were among the first to investigate whether emotional material also benefits older adults' source memory (see also Rahhal et al., 2002). In their first experiment, the authors drew on a social cover story (i.e. food served at a wedding) and presented participants with food (= items) on the left or right screen location (= source). Foods' location indicated its safety (emotional source condition) or its temperature (non-emotional source condition). Results showed that older adults' source memory improved significantly for emotionally meaningful sources (food safety) while younger adults' source memory was unaffected by source emotionality. Notably, this emotion-based benefit in older adults was strong enough to overcome the otherwise present age-related source memory deficit. That is, while the deficit occurred for the neutral source feature (i.e. temperature), it was absent for the emotionally meaningful source feature (i.e. safety).

However, later studies struggled to replicate these findings. For example, Davidson et al. (2006) used sentences (=items) spoken by voices with a neutral or emotional tone (=sources) in their study and found that older adults benefited less from emotional sources compared to younger adults. Even studies that, like May et al. (2005), specifically used threatening sources to manipulate source emotionality failed to replicate the finding that older compared to younger adults benefit more from emotional sources (Bell et al., 2013). Instead emotion-enhanced source memory occurred in younger adults at least to a similar extent, aligning with previous research focusing on younger adults only (Bell & Buchner, 2010; Smith et al., 2005; Symeonidou & Kuhlmann, 2024; but see Bisby & Burgess, 2013; Symeonidou & Kuhlmann, 2022 for opposing findings). Notably, inconsistencies in prior findings may be partially attributable to variations in item and especially source material, particularly with respect to their emotional characteristics. In our previous work (Symeonidou et al., 2022), we specifically addressed these methodological shortcomings by refining stimulus selection, allowing for tighter control over the two key emotionality dimensions – valence and arousal (Russell, 1980). This enhanced methodological rigour provided a more reliable basis for interpreting source emotionality effects across age groups. More specifically, instead of relying on un-normed emotional manipulations (as in e.g. May et al., 2005), we used normed negative, neutral, and positive background images as sources and unrelated (neutral) words as items. This created a clear distinction between items and sources and additionally allowed distinguishing between effects of positive versus negative sources. The results of this previous work suggested an emotional benefit only in younger adults' but not in older adults' source memory, again contradicting May et al. (2005). This challenges the idea of a general emotion-based effect as well as a specific positivity effect in older adults' source memory, diverging from previous findings on item memory (Reed et al., 2014). Crucially, however, in our previous work (Symeonidou et al., 2022), we had employed a relatively artificial encoding situation compared to other researchers. That is, the item-source-pairing (words with scene images) lacked socio-emotional relevance or meaning. Considering that the SST emphasizes the age-related shift in emotional and *social* priorities, this social aspect might be critical to demonstrate emotion-based benefits in older adults. It is thus worthwhile to re-investigate this question with socially relevant and more naturalistic material while retaining the methodological strengths of our previous research (e.g. matching arousal levels, clear item-source-distinction, bias-free memory measures).

Current research

As reviewed, previous studies have typically either (a) used socio-emotional material with limited control over their emotional properties or (b) used emotionally normed stimuli without social relevance. The present research integrated both: Source information consisted of normed negative, neutral, and positive scenes (following Symeonidou et al., 2022), while items consisted of faces as social cues (following Bell et al., 2013). Through their pairing with a face, the source contexts gained social relevance. Additionally, participants were instructed to imagine encountering the neutral faces within the respective scenes, further enhancing the social relevance of the sources.

Participants then rated whether the depicted face felt rather pleased or rather unpleased (pleased-unpleased ratings). Previous research has shown that such an affective encoding task encourages participants to consider sources' valence at encoding and thus facilitates emotionality effects in source memory (see Symeonidou & Kuhlmann, 2024, for a systematic investigation). Note that while such ratings might also promote valence transfer from source to item, similar to evaluative conditioning (EC; Hofmann et al., 2010), the procedure used here is not well-suited for EC to occur effectively, which is why EC is not further discussed in the following (but see Symeonidou & Kuhlmann, 2024, for a detailed discussion).

Retaining the methodological strengths of our previous work, the present research likewise systematically varied the valence of the source sceneries, while holding arousal constant at medium levels. Valence was chosen as a key variable for two main reasons: First, previous studies on emotional source memory primarily relied on valence to explain emotionality effects, often without systematically measuring or controlling for arousal (e.g. Arnold et al., 2021; Bell & Buchner, 2010; Davidson et al., 2006). Thus, it remains unclear in many studies whether observed effects stem from extreme valence, high arousal, or an interaction of both. By carefully matching arousal levels across conditions, the present study eliminates this valence-arousal confound, allowing for a clearer interpretation of valence effects (see also Symeonidou & Kuhlmann, 2022, 2024). Second, and more critically, age-related differences in emotional memory – such as the positivity bias – manifest more clearly with valenced, low-arousal stimuli. This is because these effects are thought to be driven by top-down motivational processes, which can be overshadowed by the bottom-up attentional capture associated with high-arousal stimuli (Kensinger, 2008). By keeping arousal levels as low as possible, the study maximised the likelihood of detecting age-related shifts in motivational processing.

Finally, considering the growing popularity of online behavioural research, the paradigm was administered to younger and older adults both in the lab (referred to as "Lab Experiment") and online (referred to as "Online Experiment"). This aimed to confirm robustness of results across experimental settings and examine whether the setting influences core results.

In sum, the combination of carefully-selected emotional source material, socio-emotional item-source pairs, and a direct comparison across lab and online setting reflects the methodological strengths of this research and sets it apart from previous work.

Hypotheses

Hypotheses on the pleased-unpleased ratings and on source memory were based on previous studies with comparable procedures (e.g. Bell & Buchner, 2010; Symeonidou et al., 2022). That is, both younger and older adults were expected to perceive faces (= items) in positive sceneries (= sources) as most pleased, followed by faces in the neutral sceneries, and finally followed by faces in the negative sceneries. Additionally, older adults were expected to exhibit a positivity effect in their ratings, which could express itself in two possible ways (Reed et al., 2014): Older adults would either rate all faces more positively (= positivity bias) compared to younger adults; or they would rate faces paired with negative sources less negatively (= reduced negativity bias; see Symeonidou et al., 2022).

As to source memory, previous research often found better source memory for emotional compared to neutral sources in younger adults when using affective encoding instructions (Bell & Buchner, 2010; Smith et al., 2005; Symeonidou & Kuhlmann, 2024). Thus, an emotion-based source-memory benefit was expected in the younger samples. For older adults, the evidence is overall mixed; however, source-memory benefits seem to more readily emerge with socio-emotional material (Davidson et al., 2006; May et al., 2005). Given that the herein used face-scene pairings created a socio-emotional context, an emotion-based benefit was likewise expected for older adults. Regarding the occurrence of an age-related positivity effect in source memory, existing evidence is limited, making clear predictions challenging. Note, however, that testing for emotion-based benefits inherently involves examining positivity effects as well.

The Lab and Online Experiment contained the same material and procedure with the only difference that the former employed German-speaking participants in the lab while the latter employed English-speaking participants online. Considering that source and item material consisted of images rather than text, language-specific effects are unlikely. Further, since emotion-based benefits in source memory have been previously reported both in the lab (e.g. Bell & Buchner, 2010) and online (e.g. Symeonidou et al., 2022), results were expected to replicate across experiments.

Crucially note that memory data were analysed with the two-high-threshold multinomial model of source monitoring (2HTSM; Bayen et al., 1996), extended to three sources (Keefe et al., 2002). The memory parameters are thus corrected for guessing bias, providing a clear advantage over traditionally used empirical memory measures, such as conditional source identification measures (CSIMs; Bröder & Meiser, 2007; Murnane & Bayen, 1996). Analysis with CSIMs was also conducted – yielding virtually the same results for both experiments – and is reported in the online supplement for comparison.

This research was preregistered and the data from both studies are openly available on the Open Science Framework (OSF) repository (see Data Availability Statement).

Lab experiment

Method

Participants and design

Recruitment periods of both the Lab and Online Experiment overlapped for efficiency, and thus, the a priori power-analysis was optimised for both studies. To conduct an a priori power analysis with the 2HTSM (Bayen et al., 1996; Keefe et al., 2002) specific assumptions about the parameter values of the model are required. These assumptions were based on a pilot study, which was conducted with *older* adults ($N = 25$) in an *online* setting. This approach ensured a rather conservative estimation of the required N for two reasons: (1) Online studies tend to have greater variability, requiring a larger sample size to detect effects reliably (Segen et al., 2021). (2) Older (compared to younger) adults typically exhibit lower item memory, which, for a given α and β, requires more observations to detect the same difference in source memory (because source memory is conditional on item memory in the 2HTSM). Put simply, the power analysis was specifically optimised for online-recruited older adults, for whom source-memory effects are generally harder to detect. This guaranteed an overall high power for both Lab and Online Experiment.

The assumed parameter values based on the pilot data were $D = .40$, $g_{neutral} = .39$; $g_{positive} = .62$, $b = .58$, $d_{neutral} = .64$. The effect of interest referred to the difference between source memory for the emotional (negative & positive) sources versus the neutral sources (emotionality effect). A (minimum) difference of .25 was assumed based on our previous experiments (Symeonidou et al., 2022), and thus $d_{positive} = d_{negative} = .89$ entered the power analysis. The analysis yielded a desired sample size of $n = 68$ participants per age group with $\alpha = .05$ and power of $1\text{-}\beta = .80$, that is, $N = 136$ (i.e. 2*68) in total per experiment.

Ultimately, 80 younger and 84 older adults participated in the Lab Experiment. From these, 10 younger and 16 older adults had to be excluded for the

Table 1. Sample characteristics.

Measure	Lab experiment		Online experiment	
	Younger adults	Older adults	Younger adults	Older adults
Mean age	21.61 (2.43)	73.06 (6.80)	22.37 (1.71)	68.91 (3.26)
Range age	18–29	60–91	19–25	65–79
Pattern comparison	0.78 (0.11)	0.43 (0.14)	0.80 (0.10)	0.60 (0.12)
Vocabulary	0.67 (0.12)	0.84 (0.11)	0.41 (0.14)	0.67 (0.20)
Subjective nearness to death	1.54 (0.96)	2.26 (1.25)	1.79 (0.99)	2.37 (1.12)

Note. Standard deviation in brackets. Performance in the pattern comparison task (Salthouse, 1996) and vocabulary task (from Riegel, 1967, for the German sample in the Lab Experiment; from Ekstrom et al., 1979, for the English sample in the Online Experiment) refers to the proportion of correct responses among all responses. Subjective nearness to death was rated on a 5-point Likert scale (1: strongly disagree to 5: strongly agree). Age criteria were a priori set to be narrower in the online (YA: 18–25; OA: 65+) compared to the lab study (YA: 18–30; OA: 60+), see Appendix.

following, preregistered reasons: demographic or health-related criteria (8 younger and 7 older adults), 16%[1] or more missing responses in the pleased-unpleased ratings (1 younger and 7 older adults), use of only three (or less) response options in the source-monitoring test (1 younger and 2 older adults). The demographic and health-related questions are listed in the Appendix. The final lab sample consisted of 70 eligible younger adults[2] aged $M = 21.61$ years old (range = 18–29, $SD = 2.43$) and 68 older adults aged $M = 73.06$ years old (range = 60–91, $SD = 6.80$).

For a more comprehensive sample characterisation, participants' processing speed was assessed by a pattern-comparison task (by Salthouse, 1996), and vocabulary skills were measured with the German SASKA (Riegel, 1967). Mean performance across tasks and age groups is summarised in Table 1. Compared to younger adults, older adults showed lower processing speed (classifying less patterns correctly), $t(136) = -15.85$, $p < .001$, $d = 2.70$, but better vocabulary skills, $t(136) = 8.35$, $p < .001$, $d = 1.42$. Further, older adults felt closer to death compared to younger adults, $t(136) = 3.81$, $p < .001$, $d = 0.65$, in line with the SST. Thus, for all measures, age differences were as expected.

All participants were tested in laboratory rooms. Three types of scene images (see below) were used to manipulate source emotionality (negative, neutral, positive). This results in a two × three mixed design, with age (younger vs. older) varying between participants and source emotionality (negative vs. neutral vs. positive) manipulated within participants.

Material

Faces as items

Thirty-six faces were drawn from the *FACES* database (Ebner et al., 2010). Taking the own-age-bias into account (Rhodes & Anastasi, 2012), both younger and older faces (half male, half female) were selected (see Bell et al., 2013, for a similar procedure). To ensure face neutrality, only faces that were accurately categorised as neutral by at least 80% of participants were considered. Additional criteria were mean distinctiveness ratings (taken from Ebner et al., 2018) of at least 30 (on a scale from 0 = "not distinctive at all" to 100 = "very distinctive"), as well as a mean trust ratings (taken from Pehlivanoglu et al., 2023) of 45–70 (on a scale from 0 = "not at all trustworthy" to 100 = "extremely trustworthy").[3] Since the database includes two pictures of the same person (Set A and Set B), the one with the higher accuracy rating was selected if both pictures met the above criteria. If accuracy was identical, the picture with the higher distinctiveness rating was selected. This yielded 89 eligible face pictures. Older faces had lower attractiveness ratings than younger ones, so older faces rated least attractive (6 female, 7 male) and younger faces rated most attractive (13 female, 17 male) were excluded. This resulted in a final selection of 12 faces per age-gender category (see Table 2 for the mean ratings). Using the R package *anticlust* (Papenberg & Klau, 2021), four sets of 12 faces were created, balanced by accuracy, distinctiveness, trustworthiness, and attractiveness. Each set was randomly assigned across participants to the negative, positive, or neutral scenes in the study phase or served as a distractor in the test phase.

Scene images as sources

Scene images were drawn from the *Socio-Moral Image Database* (SMID; Crone et al., 2018) a recent, large databases with overall 2,941 images. The database contains ratings on a 5-point ratings scale for (inter alia) valence (1 = unpleasant/negative to 5 =

Table 2. Mean ratings of the normed faces per age-gender category.

Age-gender category	Perceived age	Accuracy	Attractiveness	Distinctiveness	Trustworthiness
Older female	68.83 (6.19)	87.67 (30.90)	39.90 (21.71)	39.04 (22.73)	57.48 (20.30)
Older male	68.52 (6.40)	87.31 (39.47)	39.47 (20.61)	36.60 (21.73)	55.03 (20.14)
Younger female	26.41 (5.24)	93.17 (49.93)	49.93 (23.59)	39.28 (23.49)	61.56 (18.36)
Younger male	29.02 (5.34)	93.58 (43.21)	43.21 (21.78)	36.61 (22.73)	53.94 (19.02)

Note. Standard deviation in brackets. Each age-gender set consisted of 12 face pictures. Face pictures were drawn from the database FACES (Ebner et al., 2010). Ratings (for each face stimulus) for perceived age and accuracy were drawn from Ebner et al. (2010); ratings for attractiveness and distinctiveness were drawn from Ebner et al. (2018); ratings for trustworthiness were drawn from Pehlivanoglu et al. (2023).

pleasant/positive), arousal (1 = calming to 5 = exciting), and morality (1 = immoral/blameworthy to 5 = moral/praiseworthy). Overall, six images were selected per valence category (i.e. 6 negative, 6 neutral, 6 positive) such that all three sets matched in terms of mean arousal and had as low arousal levels as possible (see Symeonidou et al., 2022, for a similar procedure); the negative and positive sets matched in terms of their mean absolute valence, ensuring a comparable valence strength. All sets matched on low-level features (checked post-hoc). Images with a focal face (of a person or animal) were excluded as this could interfere with the face pictures. As noted above, the social relevance of the sources was ensured by pairing them with the face items. Since all scenes were paired with faces in the same way, socio-emotional relevance was well-matched across all three source types. Mean valence and arousal ratings are listed in Table 3. Details on the selection procedure and on low-level features are reported in the online supplement.

Procedure

The experiment was built in *lab.js* (Henninger et al., 2022). Participants were tested in laboratory rooms at the University of Mannheim in age-group-homogeneous groups. Younger adults were recruited from the university's student population via the electronic SONA system. Older adults were invited from our lab-internal database via phone calls. Demographic and health-related eligibility criteria (see Appendix) were checked in a self-report demographic survey at the end of the experiment. Non-eligible participants were excluded and replaced post-hoc after completing the experiment. After providing informed consent, participants started the encoding task (Figure 1, left side). Each trial began with a 500 ms fixation cross, followed by one of the 18 scene images (6 negative, 6 neutral, 6 positive) for 750 ms. Afterwards, one of the 36 faces (18 younger- and 18 older-looking faces) was superimposed on the scene and presented with a 5-point ratings scale underneath for 6 seconds. Participants were instructed to imagine encountering each face in the respective scene and indicated how pleased the person feels using the 5-point rating scale (1: very unpleased to 5: very pleased). If they needed less than 6 seconds for their response, the pressed number turned blue to indicate that their answer was logged. Thus, the presentation time for the face-scene pairing was fixed at 6 seconds. Each scene was presented twice, once with an older- and once with a younger-looking face. The presentation order was random with the constraint of maximum three successive same-valence sceneries (e.g. maximum 3 negative sceneries in a row). Across all trials, each participant was presented with an equal number of young and

Table 3. Means and standard deviations of valence and arousal ratings per valence category.

Emotionality	Scene images	Valence	Absolute Valence	Arousal
Negative	"b13_p223_5", "b2_p23_19", "b4_p49_8", "b11_p164_14", "b999_p492_14", "b999_p497_1"	1.42 (0.69)	1.58 (0.69)	3.11 (1.24)
Positive	"b10_p140_20", "b14_p256_17", "b10_p146_10", "b15_p295_13", "b11_p171_14", "b13_p236_8"	4.57 (0.63)	1.57 (0.63)	2.95 (1.43)
Neutral	"b14_p255_12", "b13_p221_7", "b999_p482_3", "b6_p89_4", "b999_p493_10", "b1_p3_11"	3.00 (0.90)	0.00 (0.90)	2.89 (1.21)

Note. Standard deviation in brackets. Images were drawn from the Socio-Moral Image Database (SMID; Crone et al., 2018). The labels of the images are reported under column "Scene images". Images were selected such that the negative and positive images matched on absolute valence, and all three image categories matched on arousal level. Absolute valence refers to the strength of valence, independent of direction (i.e. negative vs. positive), and is computed by subtracting the scale midpoint (3) from the valence rating.

old faces (as well as male and female faces), controlling for own-age (and own-gender) bias. Participants were not informed about any upcoming memory test; thus, item and source learning were incidental. After the study phase, participants completed a pattern-comparison task (Salthouse, 1996) for three minutes, deciding if two patterns of lines presented side-by-side were the same (key 1) or different (key 0) as quickly as possible. This task consisted of two 30-second blocks, corresponding to the paper-based version. These blocks were scored for processing speed. If participants finished in less than three minutes, a third block repeated items from the preceding blocks to fill the time but was not scored. In the subsequent test phase (Figure 1, right side), all 36 studied faces plus 12 new distractor faces were presented. Below the faces, one negative, one neutral and one positive scene were presented side-by-side with a "new" option printed at the centre bottom. The mapping of valence category (negative, neutral, positive) to screen position (left, centre, right) was randomly determined for each participant anew and kept constant across test trials. For old faces, one of the three sceneries was correct, whereas the other two were sceneries from the remaining two valence categories, which belonged to other faces. For new faces, all three sceneries were sceneries that belonged to other studied faces. Thus, each scene appeared 8 times in the test phase: twice as the correct source option (once with an older- and once with a younger-looking face), four times as the incorrect source option (twice with an older- and twice with a younger-looking face), and twice with a distractor face (once with an older- and once with a younger-looking distractor face). For each face, participants self-paced chose one of the three sceneries (negative, neutral, positive) or decided that the face was new using their keyboard. They then completed the 20-item German vocabulary task *SASKA* (Riegel, 1967). Each item was presented consecutively with five response options below, labelled 1–5. These options consisted of words or brief phrases, and participants selected the option that best matched the meaning of the target word by pressing the corresponding number key. The task was self-paced and ended once all 20 items had been addressed. Afterward, participants answered demographic and health questions and reported any issues with stimulus presentation or instructions, as well as whether they used any assistance during the task. They also rated their agreement with the statement "I have the feeling that my time is coming to an end" (Lang, 2000, p. 162) on a 5-point rating scale ranging from (1) strongly disagree to (5) strongly agree, which measured their time horizon. Finally, participants provided

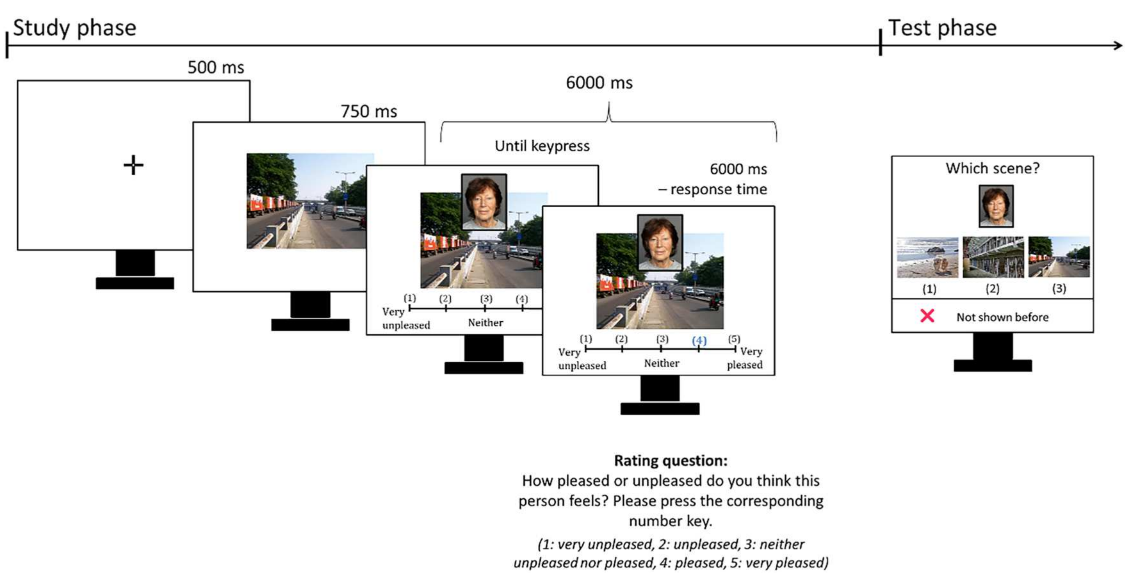

Figure 1. Illustration of the experimental flow (study & test phase).

Note. Illustrated is one trial of the study phase and one trial of the test phase. The procedure was the same in the Lab and Online Experiment with the rating scale phrased in German versus English, respectively. In both experiments, the study and test phase were separated by a three-minute interval, which was filled with a pattern-comparison task (not depicted here).

general feedback if desired and were debriefed about the study's research aim.

Results and discussion

Alpha level was set to α = .05 for all analyses.

Pleased-unpleased ratings

Figure 2 (left side) shows the mean pleased-unpleased ratings of participants dependent on source emotionality. A 2 (age group) × 3 (source emotionality) mixed ANOVA revealed a main effect of age, $F(1, 136) = 7.48$, $p = .007$, $\eta_p^2 = .05$: Older adults overall rated the faces as more pleased compared to younger adults, which replicated the positivity bias observed in previous experiments (Symeonidou et al., 2022). Furthermore, a main effect of source emotionality[4] occurred, $F(1.32,\ 180.08) = 137.05$, $p < .001$, $\eta_p^2 = .50$, but no age group × source emotionality interaction, $F(1.32,\ 180.08) = 0.55$, $p = .509$. Following up on the main effect of source emotionality, Bonferroni-Holm adjusted pairwise comparisons showed that, as expected, faces in the positive scenes were rated as more pleased compared to faces in the neutral scenes, $t(136) = 9.11$, $p < .001$, $d = 0.63$, and faces in the negative scenes, $t(136) = 12.62$, $p < .001$, $d = 1.41$. Similarly, faces in the neutral scenes, were rated as more pleased than faces in the negative scenes, $t(136) = 11.51$, $p < .001$, $d = 0.79$. This indicates that both, younger and older adults, considered scene's emotionality when rating the neutral faces, replicating previous research (Bell & Buchner, 2010; Symeonidou et al., 2022).

Source memory

As mentioned above, the 2HTSM (Bayen et al., 1996), extended to three sources (Keefe et al., 2002), was used to analyse memory data (see Figure 3) and derive bias-free memory measures. The 2HTSM assumes that three general processes contribute to performance in source-monitoring tasks: item recognition (parameter D), source memory (parameter d), and guessing processes (item old/new guessing, parameter b; source guessing, parameter a/g). Thus, it allows disentangling the contribution of memory processes versus guessing bias to the observed responses, providing bias-corrected memory measures. More specifically, the following parameters were estimated for each age group: Parameter D measures the probability of item memory, that is, memory for the faces (assumed to be equal across each source-valence category/ type of scene). Parameter d measures the probability of source memory, that is, correctly remembering the original source (i.e. scene) of the recognised face ($d_{negative}$ for the negative sources, $d_{positive}$ for the positive sources, and $d_{neutral}$ for the neutral sources). If the source cannot be remembered (e.g. 1-$d_{negative}$), guessing processes come into play. Specifically, parameter $g_{positive}$ measures the probability to guess the positive source. If 1-$g_{positive}$ (positive source is not guessed), parameter $g_{negative}$ measures the probability to guess the negative source and probability 1-$g_{negative}$ the probability to guess the neutral source. If item memory fails (1-D), participants' answers are entirely based on guessing processes: With probability b, participants guess that a face was previously presented in the study phase (i.e. is "old"), followed by guessing the

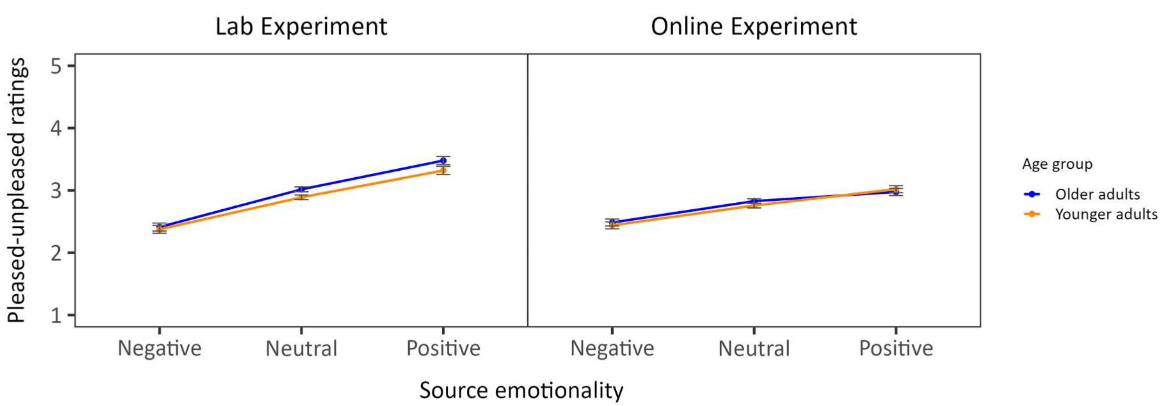

Figure 2. Pleased-unpleased ratings for the lab experiment and online experiment.

Note. Mean pleased-unpleased ratings for younger and older adults in the Lab (left-hand plot) and Online Experiment (right-hand plot) as a function of source emotionality. Error bars indicate one standard error of the mean. The scale ranged from 1: very unpleased to 5: very pleased.

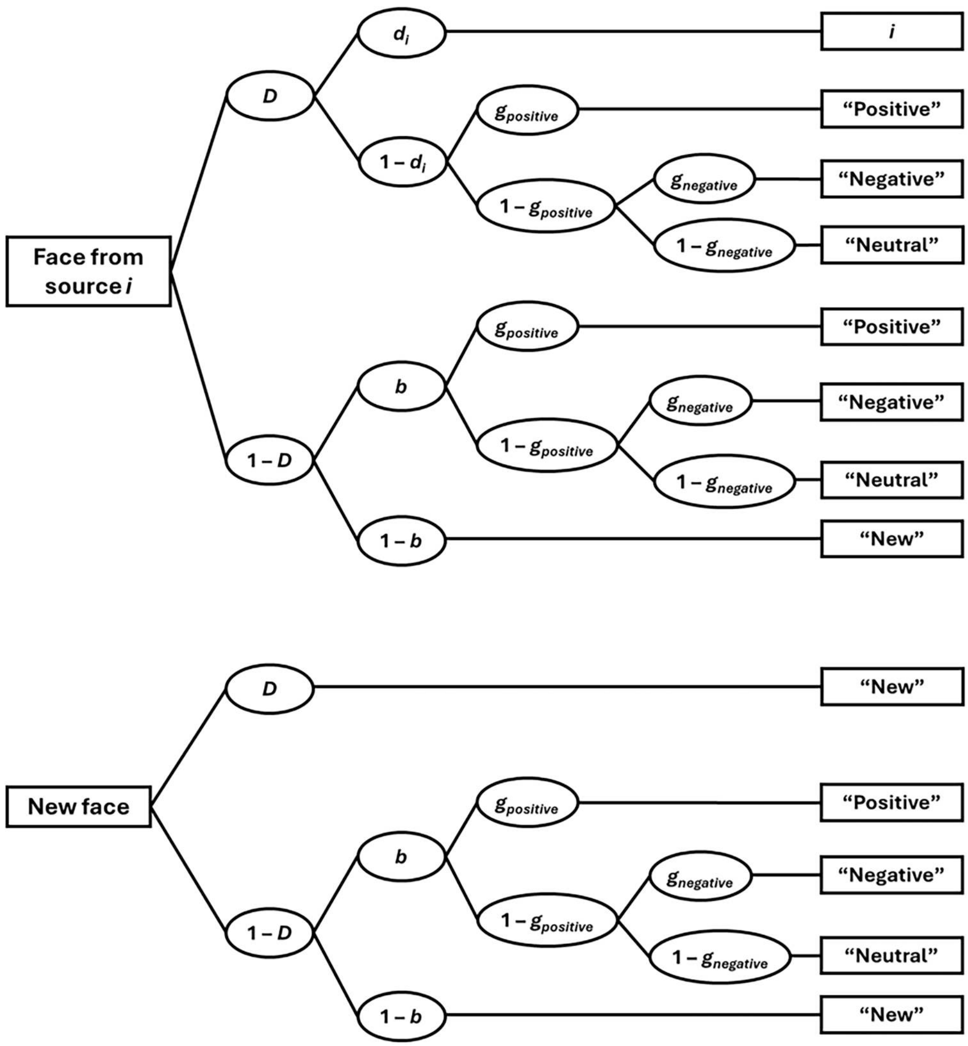

Figure 3. Graphical representation of the two-high-threshold multinomial model of source monitoring (2HTSM) for three sources.

Note. The figure shows submodel 5d of the 2HTSM for studied faces (upper tree) and for new faces (lower tree), extended to three sources. *i* denotes the emotionality of the source with which the item was originally paired, $i \in$ {negative, neutral, positive}. Boxes on the right represent participants' responses in the source-memory test. *D* = probability of recognising a previously presented face as old (equated across the negative, neutral, and positive sources) or new; d_i = probability of correctly recalling the source of a recognised face; *b* = probability of guessing that a face was previously presented; $g_{positive}$ = probability of guessing the positive source for a recognised or unrecognised face (.33 if unbiased); $g_{negative}$ = probability of guessing the negative (vs. neutral source) for a recognised or unrecognised face if the positive source was not guessed (.50 if unbiased). Adapted from "Source monitoring deficits for self-generated stimuli in schizophrenia: Multinomial modelling of data from three sources", by Keefe et al. (2002, p. 63).

associated source as either the positive ($g_{positive}$) or negative ($g_{negative}$) or neutral source (1-$g_{negative}$). With the complementary probability 1-*b*, participants guess that the face is new. Parameters were estimated based on the aggregated observed response frequencies and model fit was assessed using maximum likelihood (ML) estimation methods, as implemented in the software multiTree (Moshagen, 2010). Model fit and parameter estimates for both age groups and both experiments are reported in Table 4. Parameters were also estimated based on a Bayesian-hierarchical approach (latent-trait approach, Klauer, 2010; using the R package *TreeBUGS*, Heck et al., 2018). This method provides person-specific parameters and enables exploring their association with covariates, though it is less suited for between-group comparisons (cf., Chechile, 2009). These estimates, which closely matched those from the aggregated approach, are reported in the online supplement along with correlations between memory parameters

Table 4. Parameter estimates and model fit of the two-high-threshold multinomial model of source monitoring, extended to three sources for both age groups and experiments.

Age Group	Model Fit $G^2(5)$	Parameter estimates D	$d_{negative}$	$d_{neutral}$	$d_{positive}$	b	$g_{positive}$	$g_{negative}$
Lab Experiment								
Older adults	4.90, p = .429	.35 [.31; .39]	.52 [.37; .67]	.35 [.22; .49]	.55 [.41; .70]	.49 [.46; .52]	.34 [.31; .37]	.55 [.51; .59]
Younger adults	4.21, p = .520	.53 [.50; .57]	.61 [.52; .70]	.65 [.56; .74]	.67 [.58; .76]	.38 [.34; .41]	.33 [.29; .36]	.50 [.45; .55]
Online Experiment								
Older adults	4.98, p = .418	.37 [.34; .41]	.35 [.22; .49]	.20 [.07; .33]	.21 [.07; .35]	.49 [.46; .52]	.36 [.33; .39]	.53 [.49; .57]
Younger adults	4.97, p = .420	.36 [.32; .40]	.50 [.36; .65]	.52 [.38; .66]	.53 [.39; .68]	.52 [.49; .54]	.32 [.29; .35]	.52 [.48; .56]

Note. Brackets indicate 95% confidence intervals. D = probability of recognising a previously presented face as old (equated across the negative, neutral, and positive sources) or new; d_i = probability of correctly recalling the i = negative, neutral, or positive source of a recognised face; b = probability of guessing that a face was previously presented; $g_{positive}$ = probability of guessing the positive source for a recognised or unrecognised face (.33 if unbiased); $g_{negative}$ = probability of guessing the negative (vs. neutral source) for a recognised or unrecognised face if the positive source was not guessed (.50 if unbiased).

and participants' pleased-unpleased ratings (and other covariates).

Note that within the MPT model framework, differences between parameters are tested by model fit comparisons; thus, conducting a classical ANOVA is not possible. More specifically, to test whether source memory is enhanced for emotional compared to neutral sources, the respective source memory parameters are equated ($d_{negative} = d_{neutral}$ and $d_{positive} = d_{neutral}$, for each age group). If source memory is better for the emotional compared to neutral sources, the equality restrictions lead to a significant drop in model fit. The same logic was applied to all other difference tests reported below.

As evident in Figure 4, left plot, younger adults' source memory did not differ dependent on source emotionality, $\Delta G^2(1) = 0.07$, $p = .788$, for the positive-neutral comparison, $\Delta G^2(1) = 0.40$, $p = .527$, for the negative-neutral comparison, and $\Delta G^2(1) = 0.82$, $p = .365$, for the positive-negative comparison. Thus, different than hypothesised, younger adults' source memory was not enhanced for emotional compared to neutral sources in the lab study. In contrast, an emotion-based benefit occurred in older adults: Both, positive and negative sources were descriptively better remembered compared to neutral sources but this benefit was only statistically significant for the positive sources, $\Delta G^2(1) = 3.84$, $p = .050$, not for the negative sources, $\Delta G^2(1) = 2.42$, $p = .120$. The positive-negative difference was also not significant, $\Delta G^2(1) = 0.11$, $p = .738$. Put simply, older adults remembered the positive sources best and the neutral sources least, with the negative sources falling in between. To more directly test for emotionality effects, $d_{negative}$ and $d_{positive}$ were equated, providing a joint estimate of source memory for the emotionally valenced sources ($d_{emotional}$; see Symeonidou et al., 2022, for a similar procedure). This new

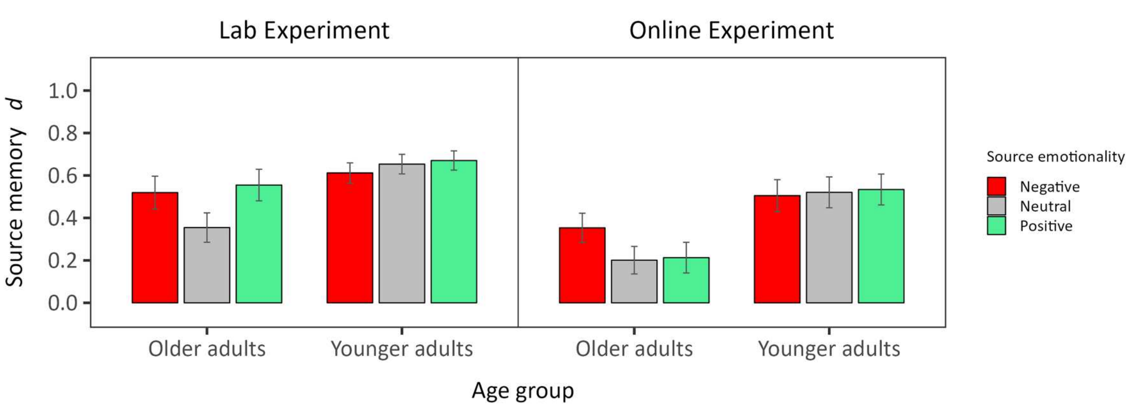

Figure 4. Source memory for negative, neutral, and positive sources for both age groups and experiments.

Note. The figure shows source memory (measured by parameter d of the 2HTSM; Keefe et al., 2002) for negative, neutral, and positive sources, separate for younger and older adults in the Lab (left-hand plot) and Online Experiment (right-hand plot), respectively. Error bars indicate one standard error of estimate.

parameter was then contrasted against source memory for neutral sources ($d_{neutral}$), again yielding a significant difference, $\Delta G^2(1) = 4.39$, $p = .036$. Thus, older adults' source memory was overall enhanced for the emotional compared to the neutral sources.

Next, to test for age effects in source memory, *d* parameters of the same source emotionality were equalised across age groups (i.e. $d^{YA}_{negative} = d^{OA}_{negative}$; $d^{YA}_{neutral} = d^{OA}_{neutral}$; $d^{YA}_{positive} = d^{OA}_{positive}$). These comparisons revealed that the age-related source memory deficit occurred only for the neutral but not for the emotional sources: While younger adults outperformed older adults in remembering neutral sources, $\Delta G^2(1) = 12.37$, $p < .001$, their performance did not differ significantly for positive sources, $\Delta G^2(1) = 1.73$, $p = .188$, and negative sources, $\Delta G^2(1) = 1.03$, $p = .310$, respectively. Put simply, the age-related source memory deficit was considerably mitigated for the emotional sources. In fact, this replicates the results reported in May et al. (2005) and contradicts our past findings (Symeonidou et al., 2022).

Furthermore, the model-based analysis was supported by the traditional CSIM analysis, reported in the online supplement, where result patterns largely replicate for both experiments with only a few exceptions. This underscores the overall robustness of the findings.

Item memory

Note that the above-described model version of the 2HTSM assumes equal item memory (i.e. recognition of the neutral faces) across the three source types. The good fit of this model to the data thus implies that item memory did not differ across sources, that is, faces paired with negative, neutral, and positive sources were recognised equally well (see Bell & Buchner, 2010; Symeonidou & Kuhlmann, 2024, for similar results). Hence, there was only one *D* parameter per age group.

As the focus of this research was on source memory, no specific hypotheses about age effects in item memory were preregistered. There is however good evidence of an age-related deficit in item memory, albeit smaller in size compared to the deficit in source memory (see Old & Naveh-Benjamin, 2008). Indeed, age-group comparisons revealed that older adults had worse item memory compared to younger adults, $\Delta G^2(1) = 54.05$, $p < .001$, in line with previous research.

To summarise, results on the pleased-unpleased ratings revealed that both younger and older adults incorporated sources' valence into their ratings of the neutral faces. Older adults additionally exhibited a positivity bias, rating all faces as more pleased compared to younger adults. Importantly, this emotionality effect also showed in older adults' source memory, such that emotional (and especially positive) sources were remembered better than neutral ones. In contrast, however, younger adults did not benefit from the emotional sources.

Online experiment

To check whether findings replicate in an online-based setting with an English-speaking sample, the same experiment was conducted via the online recruitment platform Prolific. Prolific was chosen as preferred crowdsourcing platform due to its demonstrated superiority over other platforms (e.g. MTurk) in terms of data quality and participants' naivety (Peer et al., 2022). Reliance on an English-speaking sample was unavoidable, as there are very few (i.e. less than 25) German-speaking older adults available for recruitment on Prolific. However, since all to-be-encoded material consisted of images rather than text, language-specific effects were unlikely. The online setting also allowed testing and demonstrating that online studies are a valid alternative to lab studies. Due to its convenience, time, and cost-efficiency, online behavioural research has considerably grown in popularity over the past few decades, even for complex cognitive experiments (Uittenhove et al., 2023), including those involving emotional material and older adults (e.g. Prete et al., 2024) . The development of user-friendly online programming software (e.g. *Gorilla*, Anwyl-Irvine et al., 2020; *lab.js*, Henninger et al., 2022) and crowdsourcing platforms (e.g. MTurk and Prolific) accelerated this trend. However, despite systematic investigations affirming their reliability and validity (Peer et al., 2022), concerns remain about lower data quality (Finley & Penningroth, 2015) and higher risks of self-selection bias, especially in older populations (i.e. non-representative "super agers" with high technical literacy; Dennis et al., 2021). It is thus important to promote the direct replication of studies across experimental environments, as done in the present research. This approach was planned a priori and methodological differences across experiments were minimised to ensure robust and comparable results.

Method

Participants and design

The desired sample size was $n = 68$ participants per age group, thus $N = 136$ (i.e. 2*68) in total (see Lab Experiment). From the 140 online participations (69 younger, 71 older), one younger participant was excluded due to health-related issues, one older participant due to missing responses in the ratings and another two older adults due to not using all four response options in the source-monitoring test. The final sample thus consisted of 68 eligible younger adults aged $M = 22.37$ years old (range = 19–25, $SD = 1.71$) and 68 older adults aged $M = 68.91$ years old (range = 65–79, $SD = 3.26$). Participants' processing speed was again measured with the pattern-comparison task of Salthouse (1996) and vocabulary skills were assessed with an English vocabulary task (Ekstrom et al., 1979). As expected, older adults showed lower processing speed, $t(134) = -10.48$, $p < .001$, $d = 1.80$, and better vocabulary skills, $t(134) = 8.86$, $p < .001$, $d = 1.52$, compared to younger adults, indicating that the online samples produced valid, age-appropriate data. Note that the absolute vocabulary scores are not comparable across experiments because different tasks were used for the German- (Riegel, 1967) versus English-speaking sample (Ekstrom et al., 1979). However, the mean scores are fairly comparable to other German- versus English-speaking samples of the same age from previous research (e.g. Symeonidou et al., 2022). Finally, similar to the Lab Experiment, older adults felt closer to death than younger adults, $t(134) = 3.17$, $p = .002$, $d = 0.54$. A comparison of these sample characteristics across experiments is reported in the online supplement.

Material, design, and procedure

The same faces (=items) and scene images (=sources) were used as in the Lab Experiment. Source emotionality was again manipulated within participants, resulting in a two (age: younger vs. older) × three (source emotionality: negative vs. neutral vs. positive) mixed design. The same procedure was implemented as in the lab with the following exceptions: All participants were recruited via Prolific. The same demographic and health-related eligibility criteria were applied as in the lab study with the main exception that English (not German) as native language was required (and a residency in the UK or USA). For other minor differences in eligibility screening between both experiments please refer to the Appendix. Furthermore, participants were screened in the beginning of the study with Prolific's built-in filters and with a self-report demographic survey to ensure full eligibility. Non-eligible participants were automatically excluded by the programme and thus could not continue with the remaining tasks. The experiment was hosted on the server application OpenLab (https://open-lab.online/; Shevchenko, 2022). Before starting with the actual encoding task, participants performed a scaling task to adapt the size of the scene and face pictures to their screen size and ensure that their screen was large enough to participate (i.e. browser window height of at least 14 cm ≈ 5.51 inches). Encoding, distractor and test phase were the same as in the Lab Experiment (Figure 1). Given the English-speaking sample, an English 18-item vocabulary task (part B of Ekstrom et al., 1979) was administered. The display of items and response options was as in the lab, however, an additional sixth option labelled "skip the item" was included (following Ekstrom et al.). Also, different from the German SASKA, this vocabulary task was paced and ended automatically after 4 minutes (or earlier if participants responded to all 18 items in less time). A timer and a counter on the screen showed the time already spent and the number of already answered items, respectively. The remaining procedure was as in the Lab Experiment.

Results and discussion

Of note, a direct comparison of older and younger samples *across* experiments is reported in the online supplement. Some key findings from this analysis are highlighted below to provide a more comprehensive understanding of the overall result pattern.

Pleased-unpleased ratings

A 2 (age group) × 3 (source emotionality) mixed ANOVA revealed only a main effect of source emotionality, $F(1.31,\ 175) = 60.65$, $p < .001$, $\eta_p^2 = .31$, but no main effect of age, $F(1,\ 134) = 0.30$, $p = .587$, and no interaction, $F(1.31,\ 175) = 0.77$, $p = .415$. Bonferroni–Holm adjusted pairwise comparisons of the source emotionality levels indicated that faces with positive scenes were rated overall as feeling more pleased compared to faces in neutral scenes, $t(134) = 5.39$, $p < .001$, $d = 0.34$, and faces in negative scenes, $t(134) = 8.31$, $p < .001$, $d = 0.94$, while faces in neutral scenes, were rated as more pleased than faces in negative scenes, $t(134) = 8.19$, $p < .001$, $d =$

0.58. Thus, contrary to the Lab Experiment, online-recruited older adults did not show a positivity bias. Instead, their ratings resembled those of younger adults (see Figure 2, right plot). Indeed, comparing older adults across settings (see online supplement) revealed that lab- versus online-recruited older adults gave higher ratings to faces paired with positive sources, and also higher ratings to faces paired with neutral sources. Further notably, across age groups, participants provided their ratings significantly faster online than in the lab.

Source memory

The 2HTSM (Bayen et al., 1996) was again used to analyse memory data and the same set of parameter comparisons were conducted (for the CSIM analysis, see online supplement). Replicating lab-based results, younger adults did not differ in their source memory dependent on source emotionality, $\Delta G^2(1) = 0.02$, $p = .895$, for the positive-neutral comparison, $\Delta G^2(1) = 0.02$, $p = .882$, for the negative-neutral comparison, and $\Delta G^2(1) = 0.08$, $p = .782$, for the positive-negative comparison (see Figure 4, right plot). However, different from the Lab Experiment, older adults' source memory also did not benefit from emotional sources, $\Delta G^2(1) = 0.01$, $p = .907$, for the positive-neutral comparison, $\Delta G^2(1) = 2.39$, $p = .122$, for the negative-neutral comparison, and $\Delta G^2(1) = 1.79$, $p = .181$, for the positive-negative comparison. Descriptively, source memory was somewhat enhanced for the negative sources but as reported, this difference was not statistically reliable. Thus, the emotion-based benefit ($d_{emotional} > d_{neutral}$) observed for older adults in the Lab Experiment was absent in the Online Experiment. To check whether the online study had enough power to detect the originally assumed emotion-based source memory difference of .25 (with $\alpha = .05$), a post-hoc power analysis was conducted. The analysis yielded a power of .70 for the difference, which, admittedly, is lower than intended, but not critically low.[5]

Regarding age effects, the age-group comparisons revealed that, this time, the age-related source memory deficit occurred not only for the neutral sources, $\Delta G^2(1) = 10.82$, $p = .001$, but also for the positive sources, $\Delta G^2(1) = 9.93$, $p = .002$. Interestingly, the deficit did not hold for the negative sources, $\Delta G^2(1) = 2.21$, $p = .137$. Put differently, younger adults outperformed older adults not only in remembering neutral sources – as in the Lab Experiment – but also in remembering positive sources. Thus, when comparing older adults across experiments (see online supplement), they primarily deviated in their source memory for positive sources: While these sources enhanced source memory in the lab, they did not produce the same effect online. In fact, this aligns with the above-reported divergent results on the pleased-unpleased ratings, where lab-recruited older adults showed a positivity bias whereas those online did not.

Item memory

Again, the good model fit (see Table 4) implied that item memory did not differ across sources, resulting in one D parameter per age group. The age-group comparisons revealed that older and younger adults did not differ in item memory, $\Delta G^2(1) = 0.27$, $p = .602$, again contradicting the pattern observed in the lab. Notably, comparing age groups across experiments revealed that younger adults recruited online had lower item memory than those in the lab, while older adults performed at a comparable level across settings. Further interestingly, explorative analysis[6] showed that younger adults' item memory correlated with their pleased-unpleased rating times in both experiments: the faster the rating, the lower their item memory (see online supplement). As discussed below, fast rating responses might indicate inattentiveness.

General discussion

The present research aimed to investigate whether older adults' source memory benefits from socio-emotional versus neutral sources, and whether findings are robust across experimental settings. Normed material was carefully selected based on preregistered criteria. Thus, different to previous research (e.g. Davidson et al., 2006; May et al., 2005), it was ensured that the item-source pairs had the intended (socio-) emotional character. Older and younger participants in the lab and online rated neutral faces superimposed on negative, neutral, and positive scenes by indicating how (un)pleased the face appeared. Memory data were analysed with the two-high-threshold multinomial model of source monitoring (Bayen et al., 1996; Keefe et al., 2002), controlling for guessing biases. In both the Lab and Online Experiment, pleased-unpleased ratings indicated that younger and older participants integrated scene valence into their face judgments: faces paired with positive scenes were rated most pleased, followed by neutral and positive scenes. Additionally, in the Lab Experiment, older adults exhibited a positivity bias in their face ratings, which, however, was absent in the Online Experiment. Regarding source memory, younger adults in both experiments did not benefit from emotional sources, somewhat contradicting previous findings (Bell & Buchner, 2010; Symeonidou & Kuhlmann, 2024; but see Bisby & Burgess, 2013; Symeonidou & Kuhlmann, 2022). In contrast, older adults demonstrated enhanced source memory for emotional, particularly positive, sources in the Lab Experiment, mitigating their typical deficit in source memory (see May et al., 2005, for comparable results). However, this emotional source memory benefit did not replicate in the Online Experiment, primarily due to a lacking memory-enhancing effect of positive sources. Thus, across experiments, older adults particularly differed in how they processed and remembered positive sources: Lab-recruited older adults exhibited a positivity bias that was absent among those recruited online. Finally, regarding item memory, lab-recruited younger adults outperformed their older counterparts. However, no age differences in item memory occurred online. This was because younger adults underperformed in item memory in the Online Experiment.

Emotion-based benefits for older versus younger adults

The emotion-based benefit observed in older adults in the Lab Experiment aligns with previous studies using socio-emotional material (Davidson et al., 2006; May et al., 2005), but contrasts with studies that did not incorporate such material (Symeonidou et al., 2022). This might imply that emotion-based benefits in older adults' source memory depend on the social relevance of emotional information. In the present research, socio-emotional relevance of sources was primarily induced via their pairing with face stimuli. The scenes per se did not contain any social cues. Such explicit social cues (e.g. facial expressions or gestures), however, could reinforce or alter emotionality effect in item versus source memory (cf., Stewardson et al., 2023). Future studies could thus systematically vary the presence of such social cues in emotional and non-emotional sources to test whether it is indeed the social component that makes the difference. The socio-emotional sources did not benefit older adults' source memory in the Online Experiment but note that other factors might have obscured the effect, as discussed below in detail.

More broadly, future research could more directly test whether older adults perceive socio-emotional material as more meaningful than younger adults do. This is particularly relevant given that age-related differences in emotion-enhanced memory – such as the positivity effect – are believed to stem from differences in how emotional material is appraised for its meaningfulness (Carstensen, 2006). Notably, perceived meaningfulness may be shaped not only by valence (cf., positivity effect) but also by more specific emotional features, such as safety or threat cues. In fact, adopting a more nuanced, discrete perspective on emotion, rather than relying solely on a dimensional approach, could help integrate and reconcile some of the inconsistencies in this field.

Unlike older adults, younger adults did not benefit from emotional sources in both experiments. This result somewhat deviates from previous research showing emotion-enhanced source memory in younger adults (Bell & Buchner, 2010; Smith et al., 2005), but note that findings in this area have not always been consistent (e.g. see Arnold et al., 2021; Symeonidou & Kuhlmann, 2022, for null-results). Our own research with younger adults showed that establishing emotion-enhanced source memory typically requires highlighting sources' emotional content during encoding (Symeonidou & Kuhlmann, 2024). Thus, an affective orienting task (pleased-unpleased ratings) was incorporated in the present experiments to emphasize sources' emotional character, but still no emotion-based benefits emerged for younger adults. A potential reason might be that, in addition to the affective orienting task, participants were instructed to imagine encountering the person in the respective scene. This potentially enhanced item-source integration across all scene types, not just emotional ones. In fact, in our past research with younger adults, we could show that such integrative instructions can overshadow emotion-based effects because they direct participants' attention to the item-source relationships rather than sources' (socio-)emotional character (Symeonidou & Kuhlmann, 2024).

More broadly, it is important to highlight that the emotional source scenes used in this study had low-to-moderate arousal levels – a choice intended to maximise the likelihood of observing age-related differences in emotional memory (Kensinger, 2008).

However, high arousal is often highlighted as a key factor in emotion-based memory effects in younger adults (Mather, 2007). That said, previous research has also demonstrated that high-arousal stimuli can produce inconsistent – and sometimes even detrimental – effects on associative memory, including source memory (see Chiu et al., 2013; Mather, 2007; Pereira et al., 2023, for reviews). Despite influential theoretical frameworks such as Mather's ABC theory (Mather & Sutherland, 2011), these inconsistencies remain unresolved. Isolating the effects of arousal from those of valence in source memory remains a methodological challenge, and future studies should address this gap (see Symeonidou & Kuhlmann, 2024, for a relevant discussion).

Consistent and inconsistent results across experiments

To test robustness of results and examine whether the experimental setting influences core findings, the same study, implementing identical materials and procedures, was conducted with an English-speaking sample in an online environment. Different than expected results across experiments were consistent only regarding basic cognitive measures but not regarding memory measures. More specifically, the basic measures of processing speed, vocabulary skills, and subjective nearness to death showed typical age patterns in both experiments: Younger compared to older adults showed faster processing speed, had lower vocabulary skills, and felt less close to death. This generally supports the validity of the online-collected data.

However, for the more complex (memory) measures, results did not replicate fully across experiments with two main discrepancies: Older adults primarily differed in source memory patterns while younger adults differed in item memory patterns across experiments. More specifically, for older adults, an emotion-based benefit in source memory was present in the Lab Experiment but not in the Online Experiment. For younger adults, an age-related benefit in item memory showed in the lab but not online. Of note, the experiments systematically differed not only in the test environment (lab vs. online), but also in participants' language (German vs. English). Thus, the observed discrepancies may reflect either language- and culture-related factors, or setting-related influences, such as variations in distractibility or compliance. Based on the former, one might argue that the emotional source material elicited a weaker (or even no) emotional reaction in the English-speaking versus German-speaking sample due to cultural or language-specific differences. This could explain the absence of an emotion-based benefit in the English-speaking older adults in the online study. This explanation, however, is unlikely because the source material was originally normed with an English-speaking online sample (MTurk workers with US residence and Melbourne undergraduate students; see Crone et al., 2018). Therefore, if anything, the original scene ratings should be rather valid for the English sample in the Online Experiment than the German sample in the Lab Experiment. Furthermore, since the source material consisted of images, language-sensitive or language-specific effects are improbable. Finally, language differences cannot explain the discrepancies in younger adults' item memory across experiments.

In contrast, setting-related influences can more comprehensively account for the divergent results. Online environments are often associated with higher distractibility and lower compliance compared to lab environments (Uittenhove et al., 2023), which can negatively impact participants' attention and memory. This presumably manifested in participants' pleased-unpleased ratings during encoding: The effect of the sources on participants' ratings was overall smaller in the Online Experiment ($\eta_p^2 = .31$) compared to the Lab Experiment ($\eta_p^2 = .50$), suggesting that participants attended less to the sources in the online setting. This lacking attention, in turn, potentially hampered emotion-based effects in source memory. Because the emotional benefit was observed only for older adults in the Lab Experiment, the discrepancy across experiments was particularly noticeable in this group. Lower attentiveness might also explain why younger adults showed considerably poorer item memory online compared to the lab. Additional analyses of participants' response times for the pleased-unpleased judgments, serving as an indicator of attentiveness during encoding, revealed two important result patterns: First, response times for pleased-unpleased judgments were positively correlated with younger adults' item memory. That is, those who took longer to make their judgments demonstrated higher item memory. Notably, this relationship did not emerge in older adults. Second, online-recruited participants of both age groups provided their judgments significantly faster than lab-recruited participants. Given that faster responses resulted in lower item memory for younger adults, this may explain the overall lower item memory performance in the online-recruited younger sample.

Thus, overall, attentiveness was presumably lower in the online setting, particularly in younger adults. This aligns with the notion that younger participants may be less attentive in online than in lab settings (Finley & Penningroth, 2015). Such reduced attentiveness can result from lower compliance and motivation, or from a noisier and more distracting environment, which is more likely to occur in a weakly controlled online environment than a highly controlled lab setting (Finley & Penningroth, 2015).

That said, discrepancies across settings, particularly the failure to replicate the emotionality effect in online-recruited older adults, can also be the result of reduced power. Despite a well-conceived a priori power analysis, the post-hoc power for the Online Experiment was lower than intended, though not critically low. To enhance reliability, future studies examining emotionality effects in online samples could consider increasing sample sizes.

To conclude, result discrepancies across experiments may stem from specific characteristics of the testing environment, with lower compliance and/or higher distractibility in the Online (vs. Lab) Experiment contributing to the observed differences, especially in younger adults. However, more targeted empirical investigations are needed to isolate the specific factors underlying such performance differences in future research. For older adults, the present findings challenge the notion that online-recruited samples are exceptionally capable ("super-agers"; Dennis et al., 2021). Instead, the absent age differences reported in previous online studies are more likely explained by younger adults underperforming rather than older adults overperforming.

Conclusion

The goal of this research was to investigate whether older adults' source memory is enhanced for socio-emotional versus neutral sources and whether findings are robust across experimental settings. Neutral faces were used as social items and arousal-matched negative, neutral, and positive scenes as sources, in both a lab and online setting. Regarding the two key research questions, this research offers the following answers:

(1) Do older adults benefit from socio-emotional sources?

Yes, results demonstrate that older adults show an advantage in remembering socio-emotional source information – to an extent that considerably reduced the otherwise present age-related source memory deficit.

(2) Are results robust across experimental environments?

No, they are not: Results differ between the lab and online study, despite identical materials and procedures – possibly due to greater distractibility and lower compliance in the online (vs. lab) setting. This highlights the need to acknowledge the qualities associated with different experimental settings. The environmental heterogeneity in online studies cannot match the uniformity of a highly controlled lab study. In research involving complex cognitive and emotional processing, a lab setting may therefore be more suitable. Future researchers should more deliberately consider the suitability of the experimental setting for their specific research question.

Notes

1. The original preregistered criterion was 15% of missing values. However, because some older adults did not immediately understand that they needed to use the keyboard (instead of mouse) to respond and because the threshold for outliers (2.5 SD * average number of missing responses) was 20%, the increase to 16% was reasonable.
2. The slight oversampling in lab-recruited younger adults resulted from simultaneous recruitment of participants and the post-hoc checking of eligibility.
3. Distinctiveness ratings were computed manually by averaging across male and female younger and older raters (as the authors only provided mean ratings separately for male and female younger and older participants; see Ebner et al., 2018).
4. Mauchly's Test indicated a violation of the sphericity assumption for the source emotionality factor in both the Lab and the Online Experiment. Thus, the Greenhouse-Geisser correction was applied for all ANOVA tests involving this factor.
5. Post-hoc power analysis based on the observed effect size of .20 in the Lab Experiment yielded a power of .52 (with $\alpha = .05$). Note that because the Lab and Online Experiments were conducted in parallel, the smaller-than-expected effect in the lab could not be used a priori to power the online study.
6. I thank an anonymous reviewer for suggesting this insightful analysis.

Acknowledgments

I thank Beatrice G. Kuhlmann and Carolin Streitberger for their helpful comments on the interpretation and discussion of the obtained results and previous versions of this manuscript. I further thank Miruna-Andrea Nicolaua, Amelie Scheidel, and Katrin Schreckenberger for their help in data collection for the lab study.

Data availability statement

The data supporting the results reported in this manuscript are publicly available in the Open Science Framework (OSF) repository at https://osf.io/3e7v6/. The preregistration can be accessed via https://osf.io/4gd8h/.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was funded by the "Open Science Grant 2022" of the University of Mannheim.

ORCID

Nikoletta Symeonidou http://orcid.org/0000-0002-2471-8653

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Arnold, N. R., González Cruz, H., Schellhaas, S., & Bublatzky, F. (2021). A multinomial modelling approach to face identity recognition during instructed threat. *Cognition and Emotion*, *35*(7), 1302–1319. https://doi.org/10.1080/02699931.2021.1951175

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 197–215. https://doi.org/10.1037/0278-7393.22.1.197

Bell, R., & Buchner, A. (2010). Valence modulates source memory for faces. *Memory & Cognition*, *38*(1), 29–41. https://doi.org/10.3758/MC.38.1.29

Bell, R., Giang, T., Mund, I., & Buchner, A. (2013). Memory for reputational trait information: Is social-emotional information processing less flexible in old age? *Psychology and Aging*, *28*(4), 984–995. https://doi.org/10.1037/a0034266

Bisby, J. A., & Burgess, N. (2013). Negative affect impairs associative memory but not item memory. *Learning & Memory*, *21*(1), 21–27. https://doi.org/10.1101/lm.032409.113

Bröder, A., & Meiser, T. (2007). Measuring source memory. *Zeitschrift Für Psychologie / Journal of Psychology*, *215*(1), 52–60. https://doi.org/10.1027/0044-3409.215.1.52

Carstensen, L. L. (2006). The influence of a sense of time on human development. *Science (New York, N.Y.)*, *312*(5782), 1913–1915. https://doi.org/10.1126/science.1127488

Chechile, R. A. (2009). Pooling data versus averaging model fits for some prototypical multinomial processing tree models. *Journal of Mathematical Psychology*, *53*(6), 562–576. https://doi.org/10.1016/j.jmp.2009.06.005

Chiu, Y.-C., Dolcos, F., Gonsalves, B. D., & Cohen, N. J. (2013). On opposing effects of emotion on contextual or relational memory. *Frontiers in Psychology*, *4*, 103. https://doi.org/10.3389/fpsyg.2013.00103

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, *11*(4), 464–473. https://doi.org/10.1177/1948550619875149

Crone, D. L., Bode, S., Murawski, C., & Laham, S. M. (2018). The Socio-Moral Image Database (SMID): A novel stimulus set for the study of social, moral and affective processes. *PLoS One*, *13*(1), e0190954. https://doi.org/10.1371/journal.pone.0190954

Davidson, P. S. R., McFarland, C. P., & Glisky, E. L. (2006). Effects of emotion on item and source memory in young and older adults. *Cognitive, Affective, & Behavioral Neuroscience*, *6*(4), 306–322. https://doi.org/10.3758/CABN.6.4.306

Dennis, N. A., Kuhlmann, B. G., Bennett, I., & Mills-Finnerty. (2021, January 14). *Cognitive aging online research*.

Ebner, N. C., Luedicke, J., Voelkle, M. C., Riediger, M., Lin, T., & Lindenberger, U. (2018). An adult developmental approach to perceived facial attractiveness and distinctiveness. *Frontiers in Psychology*, *9*, 561. https://doi.org/10.3389/fpsyg.2018.00561

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—a database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, *42*(1), 351–362. https://doi.org/10.3758/BRM.42.1.351

Ekstrom, R. B., French, J. W., & Harman, H. H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs*, *79*(2), 3–84.

Finley, A., & Penningroth, S. (2015). Online versus in-lab: Pros and cons of an online prospective memory experiment. *Advances in Psychology Research*, *113*, 135–162.

Glisky, E. L., Rubin, S. R., & Davidson, P. S. (2001). Source memory in older adults: An encoding or retrieval problem? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1131–1146. https://doi.org/10.1037/0278-7393.27.5.1131

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*(1), 264–284. https://doi.org/10.3758/s13428-017-0869-7

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). Lab.js: A free, open, online study builder. *Behavior Research Methods*, *54*(2), 556–573. https://doi.org/10.3758/s13428-019-01283-5

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*(3), 390–421. https://doi.org/10.1037/a0018916

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28. https://doi.org/10.1037/0033-2909.114.1.3

Keefe, R. S. E., Arnold, M. C., Bayen, U. J., McEvoy, J. P., & Wilson, W. H. (2002). Source-monitoring deficits for self-generated stimuli in schizophrenia: Multinomial modeling of data from three sources. *Schizophrenia Research*, *57*(1), 51–67. https://doi.org/10.1016/S0920-9964(01)00306-1

Kensinger, E. A. (2008). Age differences in memory for arousing and nonarousing emotional words. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *63*(1), P13–P18. https://doi.org/10.1093/geronb/63.1.P13

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Kuhlmann, B. G., & Touron, D. R. (2017). Relate it! Objective and subjective evaluation of mediator-based strategies for improving source memory in younger and older adults. *Cortex*, *91*, 25–39. https://doi.org/10.1016/j.cortex.2016.11.015

Lang, F. R. (2000). Endings and continuity of social relationships: Maximizing intrinsic benefits within personal networks when feeling near to death. *Journal of Social and Personal Relationships*, *17*(2), 155–182. https://doi.org/10.1177/0265407500172001

Mather, M. (2007). Emotional arousal and memory binding: An object-based framework. *Perspectives on Psychological Science*, *2*(1), 33–52. https://doi.org/10.1111/j.1745-6916.2007.00028.x

Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science*, *6*(2), 114–133. https://doi.org/10.1177/1745691611400234

May, C. P., Rahhal, T., Berry, E. M., & Leighton, E. A. (2005). Aging, source memory, and emotion. *Psychology and Aging*, *20*(4), 571–578. https://doi.org/10.1037/0882-7974.20.4.571

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, *135*(4), 638–677. https://doi.org/10.1037/a0015849

Moshagen, M. (2010). Multitree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42–54. https://doi.org/10.3758/BRM.42.1.42

Murnane, K., & Bayen, U. J. (1996). An evaluation of empirical measures of source identification. *Memory & Cognition*, *24*(4), 417–428. https://doi.org/10.3758/BF03200931

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1170–1187. https://doi.org/10.1037/0278-7393.26.5.1170

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging*, *23*(1), 104–118. https://doi.org/10.1037/0882-7974.23.1.104

Papenberg, M., & Klau, G. W. (2021). Using anticlustering to partition data sets into equivalent parts. *Psychological Methods*, *26*(*2*), 161–174. https://doi.org/10.1037/met0000301

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Pehlivanoglu, D., Lin, T., Lighthall, N. R., Heemskerk, A., Harber, A., Wilson, R. C., Turner, G. R., Spreng, R. N., & Ebner, N. C. (2023). Facial trustworthiness perception across the adult life span. *The Journals of Gerontology: Series B*, *78*(3), 434–444. https://doi.org/10.1093/geronb/gbac166

Pereira, D. R., Teixeira-Santos, A. C., Sampaio, A., & Pinheiro, A. P. (2023). Examining the effects of emotional valence and arousal on source memory: A meta-analysis of behavioral evidence. *Emotion*, *23*(6), 1740–1763. https://doi.org/10.1037/emo0001188

Prete, G., Palumbo, R., Ceccato, I., Di Crosta, A., La Malva, P., Sforza, V., Laeng, B., Tommasi, L., Di Domenico, A., & Mammarella, N. (2024). The effect of aging on the dual-route model of emotion processing applied to memory recognition. *Neuropsychology*, *38*(7), 679–686. https://doi.org/10.1037/neu0000972

Rahhal, T. A., May, C. P., & Hasher, L. (2002). Truth and character: Sources that older adults can remember. *Psychological Science*, *13*(2), 101–105. https://doi.org/10.1111/1467-9280.00419

Reed, A. E., Chan, L., & Mikels, J. A. (2014). Meta-analysis of the age-related positivity effect: Age differences in preferences for positive over negative information. *Psychology and Aging*, *29*(1), 1–15. https://doi.org/10.1037/a0035194

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin*, *138*(1), 146–174. https://doi.org/10.1037/a0025750

Riegel, K. F. (1967). *Der sprachliche Leistungstest SASKA* [The verbal aptitude test SASKA]. Hogrefe.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403–428. https://doi.org/10.1037/0033-295X.103.3.403

Segen, V., Avraamides, M., Slattery, T., Colombo, G., & Wiener, J. M. (2021). Comparable performance on a spatial memory task in data collected in the lab and online. *PLoS One*, *16*(11), e0259367. https://doi.org/10.1371/journal.pone.0259367

Shevchenko, Y. (2022). Open Lab: A web application for running and sharing online experiments. *Behavior Research Methods*, *54*(6), 3118–3125. https://doi.org/10.3758/s13428-021-01776-2

Smith, A. P. R., Henson, R. N. A., Rugg, M. D., & Dolan, R. J. (2005). Modulation of retrieval processing reflects accuracy of emotional source memory. *Learning & Memory*, *12*(5), 472–479. https://doi.org/10.1101/lm.84305

Stewardson, C. I., Hunsche, M. C., Wardell, V., Palombo, D. J., & Kerns, C. M. (2023). Episodic memory through a social and emotional lens. *Emotion*, *23*(4), 961–972. https://doi.org/10.1037/emo0001147

Symeonidou, N., Hassan, A., Porstein, I., & Kuhlmann, B. G. (2022). Is there an emotionality effect in older adults' source memory? *Aging, Neuropsychology, and Cognition*, *30*, 687–712. https://doi.org/10.1080/13825585.2022.2078778

Symeonidou, N., & Kuhlmann, B. G. (2022). Better memory for emotional sources? A systematic evaluation of source valence and arousal in source memory. *Cognition and Emotion*, *36*(2), 300–316. https://doi.org/10.1080/02699931.2021.2008323

Symeonidou, N., & Kuhlmann, B. G. (2024). Enhanced source memory for emotionally valenced sources: Does an affective orienting task make the difference? *Cognition and Emotion*, *38*, 508–529. https://doi.org/10.1080/02699931.2024.2309707

Uittenhove, K., Jeanneret, S., & Vergauwe, E. (2023). From lab-testing to web-testing in cognitive research: Who you test is more important than how you test. *Journal of Cognition*, *6*(1), 13. https://doi.org/10.5334/joc.259

Appendix. Demographic and health questionnaire used in the lab and the online experiment.

Table A1. English and German wording of the demographic and health-related questions used in the lab experiment versus online experiment.

General Topic	Wording of questions in German (Lab Experiment)	Wording of questions in English (Online Experiment)	Eligibility criterium
1. Age	**Bitte tragen Sie Ihr Alter ein (in Jahren):**	**What is your current age in years?**	**YA lab: 18–30** **YA online: 18–25** **OA lab: 60+** **OA online: 65+**
2. Native Language	**Ist Deutsch Ihre Muttersprache? (Translates into: *Is German your native language?)***	**What is your first language?***	**Lab: Yes; No, but I learned it before the age of 6** **Online: English**
3. Biological Sex	Was ist Ihr biologisches Geschlecht?	What is your biological sex?	—
4. Subjective Health	Wie würden Sie Ihre derzeitige Gesundheit einstufen?	Please indicate your current health status.	—
5. Head Injury	Hatten Sie jemals eine Kopfverletzung, die dazu geführt hat, dass Sie bewusstlos wurden (z. B. durch einen Sturz, Schlag auf den Kopf, Verkehrsunfall)?	Have you ever had an injury to the head that's caused you to be knocked out for a period of time (E.g. from a fall, blow to the head, road traffic accident)?*	**Lab: —** **Online: No**
6. Respiratory disease	**Leiden Sie an einer chronischen Atemwegserkrankung? (Translates into: *Do you suffer from a chronic respiratory disease?)***	**Do you suffer from any respiratory diseases, such as asthma or chronic obstructive pulmonary disease (COPD)?***	**Lab: No; Yes, I have Asthma Online: No**
7. Heart issues	Wurden Ihnen jemals von einem Arzt/ einer Ärztin mit Herzversagen diagnostiziert?	Have you ever been diagnosed with heart failure by a medical doctor?*	No
8. Mental health	**Wurde bei Ihnen innerhalb der letzten 6 Monate eine Depression, Angststörung oder eine andere psychische Erkrankung diagnostiziert oder waren Sie innerhalb der letzten 6 Monate deswegen in Behandlung? (Translates into: *Have you been diagnosed with depression, anxiety disorder, or any other mental health condition, or have you received treatment for them in the last 6 months?)***	**Do you have – or have you had – a diagnosed, on-going mental health/ illness/condition?***	No
9. Dementia	Wurde bei Ihnen durch einen Arzt eine Demenzerkrankung (z. B. vaskuläre Demenz, Alzheimer Krankheit) diagnostiziert?	Have you ever been diagnosed with mild cognitive impairment or dementia?*	No
10. Alcohol use	***Question not used because it is covered by the broader question listed last (on addiction)***	**Have you previously been in individual therapy for alcohol use?***	**Lab: —** **Online: No**
11. Antidepressants	Nehmen Sie derzeit Medikamente zur Behandlung von Symptomen wie Depressionen, Angst oder Niedergeschlagenheit ein (z.B. SSRIs)?	Are you currently taking any medication to treat symptoms of depression, anxiety or low-mood (e.g. SSRIs)?*	No
12. Diabetes	Haben Sie Diabetes?	Do you have diabetes?	—
13. Hypertension	Haben Sie Bluthochdruck (>140mmHg systolisch)?	Do you have a hypertension (high blood pressure; > 140mmHg systolic)?	Yes, but I control it with medication; No
14. Stroke	Hatten Sie jemals einen Schlaganfall?	Did you ever have a stroke?	No
15. Parkinson	Haben Sie die Parkinsonsche Krankheit?	Do you have Parkinson disease?	No
16. Chemotherapy	Wurde Sie schon einmal mit einer Chemotherapie behandelt (z.B. aufgrund einer Krebserkrankung)?	Have you ever been treated with chemotherapy (e.g. due to cancer)?	No

(*Continued*)

Table A1. Continued.

General Topic	Wording of questions in German (Lab Experiment)	Wording of questions in English (Online Experiment)	Eligibility criterium
17. Benzodiazepines	Haben Sie im letzten Monat Benzodiazepine (z.B. Lorazepam/ Tavor®, Diazepam/Valium®, Oxazepam/Adumbran®, o.ä.) eingenommen?	Have you taken benzodiazepines (e.g. Lorazepam/Tavor®, Diazepam/Valium®, Oxazepam/Adumbran®, or similar) within the last month?	No
18. Addiction	Sind oder waren Sie jemals abhängig von Alkohol, Medikamenten oder Drogen?	Do you or did you ever suffer from addiction to alcohol, medication, or drugs?	No

Note. OA = older adults, YA = younger adults. All questions were part of the self-report survey administered either at the end (lab) or at the beginning (online) of the study. If not otherwise noted, the English questions were literally translated into German. Bold-faced are questions which slightly differed in their wording or eligibility answer across experiments. With asterisk are questions that were also used as screening filters on Prolific (built-in filters) – note that these questions had to be identical in wording to the Prolific questions, because only then a mismatch in participants' answers (between this survey and Prolific's registers) clearly justified participants' exclusion. Before recruitment began, the age criterion for older adults in the online study was adjusted from the preregistered 60+ to 65+ using Prolific's screening filters. This ensured that age means per sample were better matched across settings. As younger adults on Prolific tend to be older (25–30) than typical lab-recruited students, and older adults aged 60–65 occupy study slots more quickly, these age ranges (YA: 25–30; OA: 60–65) were deliberately excluded in the Online Experiment. The head injury question was removed as eligibility question in the lab study because it was worded too broadly, including mild injuries irrelevant to cognitive aging. Notably, serious head injuries were screened separately via phone with a more diagnostic question, and affected participants were screened out, ensuring that all participants met the study's inclusion criteria.