**The role of positive and negative emotions on multiple components of episodic memory ("what", "when", "in which context") in older compared to younger adults: a pre-registered study**

Pierrick Laulan, Ulrike Rimmele

**Response to reviews**

Dear Dr. Wonnacott, dear reviewers,

Thank you very much for your conscientious analysis of the second version of our manuscript and for your comprehensive feedback. We have taken great care to consider all comments and have provided a response to further improve our work.

As in the first round of reviews, you'll find a point-by-point response to your comments and questions: reviewers' comments are in bold, our responses are in italics, and changes made in the body of the manuscript are highlighted in yellow (all relevant changes have also been made in the design table). The changes are also indicated in yellow in the manuscript itself.

We hope you will consider the manuscript eligible for acceptance after these changes have been made.

**Revision invited - Liz Wonnacott**

Thank you for submitting this revised document. I was able to secure a review from one of the original reviewers, the other reviewer was unfortunately unavailable however I have also read the paper myself and feel I have enough to make a decision.

**1) In short, I agree with the reviewer that this is a strong revision. The reviewer has raised only a few, fairly minor, points. I have one remaining concern about the power analyses: In the new manuscript, you now have two hypotheses for each RQ- an (a) hypothesis where you look at the effect of valence in each of the three groups separately and a (b) hypothesis where you will look for the interaction between valence and age. However, it looks like at present your power analyses only address power for (b) – the interaction. Can you also include analyses to check that you will have power for looking (a) i.e. test that you can detect effect of valance with the planned individual group size?**

*Response: Thank you very much for pointing out this important element regarding the accuracy of our study in testing each of the six hypotheses mentioned in the manuscript. We performed sensitivity analyses to determine what effect size we would be able to detect for the valence factor independently in younger and older adults with a power of 0.981 (i.e., the square root of 0.964, the target power for each of our hypotheses in order to obtain a power of 0.80 over our entire study). These analyses performed on MorePower have shown that with a repeated measures ANOVA (RM-ANOVA), within an age group (i.e., younger or older adults), we are able to show a valence effect of size* $\eta^2 = 0.08$ [1] *with an alpha of .05 and a power of 0.981. Thus, under current conditions, the precision of our study is not satisfactory, and we have decided to increase our maximum sample size. To do this, we used a cost-based approach (see Lakens, 2022), considering the difficulty of recruiting a large number of older participants within a limited timeframe (i.e., between 1 and 1.5 years). We determined that targeting a maximum sample size of 300 participants (150 younger adults and 150 older adults) provides a satisfactory cost-benefit ratio. Indeed, sensitivity analyses revealed that :*

*- With 150 young adults and 150 older adults, using a RM-ANOVA, we are able to detect a within-subjects effect of valence in each age group (i.e., 3 [Valence: positive vs. neutral vs. negative] in young adults and in older adults; see hypotheses 1a, 2a and 3a) of size* $\eta^2 = 0.06$, *with an alpha of .05 and a power of 0.964 .*

*- With 150 young adults and 150 older adults, using a RM-ANOVA, we are able to detect a within-between interaction effect between valence and age (i.e., 3 [Valence: positive vs. neutral vs. negative] x 2 [Age: young vs. older adults]; see hypotheses 1b, 2b and 3b) of size* $\eta^2 = 0.03$, *with an alpha of .05 and a power of 0.964.*

*In the light of these new analyses and new information, we have completely reformulated the power analyses presented in our manuscript.*

---

[1] Please note that, when referring to effect sizes in the manuscript, we now use eta squared as measures and not Cohen's d to avoid certain approximations regarding our 3 (valence) x 2 (age) experimental design. This decision led us to repeat the sensitivity analyses presented on p. 12 of our manuscript for the studies by Waring & Kensinger (2009), Nashiro & Mather (2010), and Ceccato et al. (2022). Regarding the latter study, we have clarified how our analyses differ from those presented by the authors in their article. See p.17: In their article, the authors reported that their experimental design allowed them to detect effects of size $\eta^2 \geq 0.14$. Their power analysis was performed using G*Power (Faul et al., 2009) with the "as in Cohen" effect size specification. This calculation differs from that in MorePower (version 6.0.4; Campbell & Thompson, 2012), which corresponds to the "as in SPSS" option available in G*Power. For greater clarity and consistency in our article, we have recalculated the minimum detectable effect size in the study by Ceccato et al. (2022) using the "as in SPSS" specification in G*Power.

See p. 16-17: "Given the complexity of accessing a population of physically and psychologically healthy older adults, a problem that has been particularly evident in our recent research, especially for studies requiring a time commitment of more than 2 hours, we adopted a pragmatic approach to justifying our experimental sample size (see Lakens, 2022). This approach is guided by a recognition of the limitations inherent in our research context and a desire to maximize the scientific value of our study despite these challenges. To complete our data collection over a period of 1 to 1.5 years, it was decided to target a maximum of 150 younger and 150 older adults in our study. A first sensitivity analysis conducted with MorePower (version 6.0.4; Campbell & Thompson, 2012) revealed that, based on our target experimental sample of 150 young adults and 150 older adults, using a repeated-measures ANOVA (RM-ANOVA), we are able to detect a within-subjects effect of valence in each age group (i.e., 3 [Valence: positive *vs*. neutral *vs*. negative] in young adults and in older adults; see hypotheses 1a, 2a and 3a) of size $\eta^2 = 0.06$, with an alpha of .05 and a power of 0.964[2,3]. A second sensitivity analysis conducted under the same conditions and with the same parameters as the first revealed that we are able to detect a within-between interaction effect between valence and age (i.e., 3 [Valence: positive *vs*. neutral *vs*. negative] x 2 [Age: young *vs*. older adults]; see hypotheses 1b, 2b and 3b) of size $\eta^2 = 0.03$, with an alpha of .05 and a power of 0.964."

*We have also updated the details of our sequential analyses to reflect the new maximum sample size in our study:*

See p.17: "Thus, if the hypothesis tests all return a $p < .003$ after the first interim analyses (i.e., after 150 participants, including 75 young adults and 75 older adults), data collection will be interrupted; if the hypothesis tests all return a $p < .018$ or a $p > .298$ after the second interim analyses (i.e., after 226 participants, including 113 young adults and 113 older adults), data collection will be interrupted."

*Finally, the abstract et the design table have been modified in line with the changes described above and the references below have been removed from the bibliography:*

[2] Given that we aim to test 6 independent hypotheses divided into 3 blocks (respectively hypotheses 1a and 1b, 2a and 2b, 3a and 3b) with a power of 0.80, we have chosen a targeted power of 0.964 for each of the hypotheses, i.e. the sixth root of 0.80.

[3] We decided to base our power analysis on an RM-ANOVA because, to our knowledge, there are no software packages that allow us to directly implement a power analysis for generalized estimating equations (GEE; see Kal et al., 2022). Also, with GEE, a smaller number of participants is required in order to obtain satisfactory power compared to an RM-ANOVA (Ma et al., 2012), so our power analysis is conservative.

Brysbaert, M. (2019). How Many Participants Do We Have to Include in Properly Powered Experiments? A Tutorial of Power Analysis with Reference Tables. *Journal of Cognition*, *2*(1), 16. https://doi.org/10.5334/joc.72

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, *1*(1), 9. https://doi.org/10.5334/joc.10

Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology*, *15*(3), e2000797. https://doi.org/10.1371/journal.pbio.2000797

**2) Also- and this is probably me being a bit pedantic- I would word the 1a hypotheses each as two hypotheses "[effect X holds] for younger adults" and "[effect X holds] for older adults" rather than"- [effect X holds] for both younger and older adults" (i.e. acknowledging the fact that one hypothesis could be confirmed and nor the other.)**

*Response: Thank you very much for emphasizing the need to clarify the hypotheses. We have therefore reworded hypotheses 1a, 2a and 3a to make it clear that these hypotheses comprise two distinct hypotheses, one in younger adults and one in older adults. We have also reworded the presentation of our objective in the Power analyses and sample size estimation section.*

See p. 15:

***Hypothesis 1: Effect of age and emotion on item memory***

(a) Positive and negative images are better remembered than neutral images in young adults. Positive and negative images are better remembered than neutral images in older adults (e.g., Denburg et al., 2003).

[…]

***Hypothesis 2: Effect of age and emotion on temporal memory***

(a) Positive and negative images are positioned more accurately in time than neutral images in young adults (e.g., D'Argembeau & Van der Linden, 2005). Positive and negative images are positioned more accurately in time than neutral images in older adults (Nashiro & Mather, 2010; Palumbo et al., 2018).

[…]

***Hypothesis 3: Effect of age and emotion on memory for the association between an item and its extrinsic context***

(a) Young adults remember less well in which videos the positive and negative images had been presented compared to in which videos the neutral images had been presented, i.e., the associations between positive and negative images and videos are less well remembered than associations between neutral images and videos in young adults (MacKenzie et al., 2015). Older adults remember less well in which videos the positive and negative images had been presented compared to in which videos the neutral images had been presented, i.e., the associations between positive and negative images and videos are less well remembered than associations between neutral images and videos in older adults (Nashiro & Mather, 2010).

See p. 16 : "We are interested in testing the existence of a main effect of valence independently in younger and older adults (hypotheses 1a, 2a, and 3a) and of a valence x age interaction effect (hypotheses 1b, 2b, and 3b) in (1) an item memory task, (2) a temporal judgment task, and (3) an associative memory task."

**Reviewer 2**

The authors did a tremendous job revising their RR. This will be an interesting study.

I have only a few remaining comments that I hope will be helpful to the authors in sharpening some of their ideas:

**1) Based on the authors' response to one of my initial queries, it is now clearer that the authors are interested in using low arousal content (their decision to do so in relation to aging makes sense). Yet, I just want to caution that it is my understanding that some theoretical models in the literature pertain to arousal and not valence, e.g., ABC: Arousal-Biased-Competition model from Mather's group. The authors do not explicitly reference ABC but the authors do cite "Mather, 2007, which is about "Emotional Arousal and Memory Binding." I think the field is still trying to understand arousal versus valence effects (with some authors emphasizing arousal and others valence). Thus, I think it is crucial that the authors take great caution in applying theories of "emotion" to their paradigm if some of such theories are rooted in arousal, whereas their manipulation is chiefly one of valence. Indeed, you almost miss the low arousal**

**decision in the methods and I might suggest making this decision more salient in the introduction, but if not, it would be good that this is discussed carefully in the discussion later. Thus, I am not suggesting a large change to the introduction but a subtle "pulling out" of this valence / arousal consideration or at least exercising a bit more caution so the reader is well aware of the nuances at play and caveats in relation to what aspects of the literature their own hypotheses stem from. Just to give one example (but there are certainly other places this can come up), the authors state: "In view of the low level of evidence in the literature regarding the combined effects of age and emotion on the memory for intrinsic item features and extrinsic item context, we are cautious in formulating our hypotheses about temporal memory and associative memory for context extrinsic to the items." I might go farther and state "particularly for low arousal content."**

*Response: We thank the reviewer for highlighting the importance of distinguishing the mechanisms underlying valence and arousal effects when examining the joint effects of age and emotion on memory, particularly in light of our focus on valence vs. arousal effects related to the choice to select only low-arousal stimuli in our study. In line with the reviewer's suggestion, we have added in our manuscript a sentence to emphasize that our hypotheses should be viewed with caution with regard to our use of low-arousal stimuli, which differs from what has been done most of the time in the literature (e.g., Ceccato et al., 2023; Palombo et al., 2021). This distinction will be discussed in more detail in the discussion.*

See p. 14: "In view of the low level of evidence in the literature regarding the combined effects of age and emotion on the memory for intrinsic item features and extrinsic item context, we are cautious in formulating our hypotheses about temporal memory and associative memory for context extrinsic to the items. This caution is all the more justified given that we chose to use low-arousal stimuli in our study, in contrast to the majority of studies mentioned in our introduction that favored high-arousal stimuli (e.g., Ceccato et al., 2023; Nashiro & Mather, 2010, 2011)."

**2) The authors made some improvements to the way they reference different types of temporal memory paradigms but they use the term "source memory" a little inconsistently: when discussing some of the paradigms "which list/session" [which to me are source memory paradigms], the authors do not consistently reference those as such, even though they use that term elsewhere. My read of their intro and/or the literature is**

**that it is not so clear yet that "the beneficial effect of emotion on memory for temporal information seems to be robust when it concerns the moment when an event occurred" which is based on source memory findings, so I wonder if the authors want to (1) use more consistent terminology and (2) just soften this a little more. I understand that the authors believe the findings of Ceccato et al., 2022 (which did not observe an emotional memory effect in a source memory paradigm) might be due to the delay but for now this is speculation. e.g., how about: "the beneficial effect of emotion on memory for temporal information seems to be robust when it concerns the moment when an event occurred, though this is based on a small number of studies with some exceptions"**

*Response: In reviewing our manuscript, we found only one reference to "source memory" (see p. 6), so in the interest of clarity and consistency in how we present the results of the literature in our introduction, this term has been removed.*

*In addition, we have followed the reviewer's recommendation and nuanced the statement regarding the robustness of emotional effects on "when" memory.*

See p. 6: "If we focus on "when" memory, Petrucci and Palombo (2021) concluded that the beneficial effect of emotion on memory for temporal information seems to be robust, though this is based on a small number of studies with some exceptions."

**3) If I may propose, the authors should refrain from using "he/she" but instead use either "he/she/they or another variant to highlight the diversity of gender and make the language more inclusive (n.b., I assume the authors plan to ensure gender is matched across young and old and may want to include that in the methods). Apologies for not pointing out earlier.**

*Response: We thank the reviewer for this suggestion. We have replaced all occurrences of "s/he" with "he/she/they" and all occurrences of "his/her" with "his/her/them" (see p. 25). We have also added the information that younger and older adults will be matched by gender in the Participants section of the Method.*

See p. 18: "We will also ensure that young adults and older adults are matched by gender."

**4) Small typo: itemcontext should be "item context"**

*Response: Thank you for pointing out this error, the correction has been made.*

See p. 7: “In addition, these effects of emotion on the memory for extrinsic item context in young adults have been demonstrated mainly for negative *vs.* neutral items.”