Thank you for your submission titled “Not fleeting but lasting: Limited influence of aging on implicit adaptative motor learning and its short-term retention”. The paper addresses the question of whether aging affects the implicit component of learning . In my own reading of the article (and echoed by the comments from the two reviewers), this a very well-done study with several strengths– the paper addresses an important question and uses a well-established approach to addressing the question. In addition, the methods and analysis are also rigorously done, and presentation of the results is transparent (with robustness checks also being provided).

However, the reviewers have raised a few concerns that need addressing before I can provide a recommendation. As you will see, both reviewers have provided thoughtful reviews and suggestions – so I hope the authors will address these concerns. Both reviewers raise important theoretical and methodological clarifications that will help improve the clarity of the manuscript. I will simply highlight what I see as a main concern here (mentioned by Reviewer 2).

The main concern is the framing of the paper as a conceptual replication. As highlighted in several discussions on the replication crisis (notably by Chambers in “The 7 Deadly Sins of Psychology”), unlike a direct replication, the idea of a “conceptual” replication is subjective and often difficult to get agreement on since it is difficult to determine apriori if the parameters that were changed from the original study are critical to the original result (While the arguments against conceptual replication are typically used when a conceptual replication ‘agrees’ with the original result, it is probably fair to say that they also apply when the conceptual replication ‘fails’, as in this case). Thus, while the current study seems to a fair and rigorous test of the original question (with a section in the Discussion highlighting differences from the original study), the framing at several points in the paper (from the Abstract to the conclusion) often blurs this difference between a conceptual and a direct replication (e.g., “We failed to replicate…”). This could potentially be misleading as it could lead readers to think that this was a direct replication failure.

I would therefore suggest that ‘conceptual replication’ framing needs some re-consideration. One possibility is that the authors could choose to address this as a study of the research question on its own (while of course still comparing the results from the original study). However, if the authors want to retain this framing of a conceptual replication, it might help to be more transparent right from the outset that conceptual replications have some weaknesses (and make sure that the terms such as ‘failed to replicate’ are not used ambiguously in the manuscript)

Thank you editor for taking care of this submission. We believe that our work represents indeed a conceptual replication of the work of Trewartha and not a direct one. The Noba project (https://nobaproject.com/modules/the-replication-crisis-in-psychology) defines a conceptual replication as “A scientific attempt to copy the scientific hypothesis used in an earlier study in an effort to determine whether the results will generalize to different samples, times, or situations.”, which is precisely what we attempted to do. In other words, a conceptual replication is when a scientist tries to confirm the previous findings using a different set of specific methods that test the same idea.

Throughout the paper, we tried to emphasize that we were indeed doing a conceptual replication of the work of Trewartha and not a direct one, such as on lines 453-459 where we wrote (after having listed the many differences between our study and that of Trewartha):

“For all these reasons, our study represents a conceptual replication of the study by Trewartha et al. 2014 and not a direct/exact replication. If any of the factors identified as differences between our study and that of Trewartha is responsible for the difference in outcomes, it means that the age-related effect on spontaneous recovery, if it exists, is highly sensitive to the experimental conditions. By employing multiple methodologies, conceptual replications provide a robustness test of the findings. Our study suggests that the generalizability and robustness of the original results should be considered with caution. The age-related difference in spontaneous recovery found by Trewartha and colleagues might be true but is likely dependent on the experimental conditions."

A minor concern is that the data file in the repository seems to be one big MATLAB file which can only be used with the author's code as far as I could tell (but this code is provided). However, it would be helpful to have the deposited dataset be "independently readable" (i.e., a datafile accompanied by a codebook that explains what these variables are and how they are organized) so that a potential reader could independently use the dataset without having to rely on the author's code.

The dataset contains both one big Matlab file that can be used to reproduce all the figures and analyses from the paper and all the raw data that can be used if one wants to analyse the data further. We added one metadata file where we documented the variables from the dataset (both the columns of the AllSubjectsData.mat file but also the fields of the ANA and TRA structures.

Reviewer 1 Kevin Trewartha

Review of the article "Not fleeting but lasting: Limited influence of aging on implicit adaptative motor learning and its short-term retention" by Pauline Hermans, Koen Vandevoorde, and Jean-Jacques Orban de Xivry, submitted to PCI Health and Movement Science.

Summary

The authors present an article aimed at evaluating the effect of aging on implicit learning processes involved in sensorimotor adaptation, with an attempt to "conceptually replicate" a finding from a previous study (Trewartha et al., 2014). Trewartha and colleagues observed an age difference in the spontaneous recovery of a previously learned movement adaptation after it is extinguished through a short period of de-adaptation. This finding is at odds with more recent evidence suggesting that implicit learning and retention is preserved in aging. A key novel contribution of this paper is a direct measurement of implicit adaptation in a force field paradigm, and an assessment of the relationship between implicit learning and spontaneous recovery in older adults. A second aim was a conceptual replication of the previously observed age difference in implicit short-term retention in the form of spontaneous recovery. The results are interesting, and the some of the conclusions drawn are supported by the analyses presented in the paper. The observations that A) implicit learning levels measured during adaptation are similar between younger and older adults, and B) that implicit learning levels correlated with spontaneous recovery, were especially interesting and novel findings. The authors also did not observe a reduced spontaneous recovery, in contrast to the previously reported finding. The paper would generally make a nice contribution to the literature on age-related changes in sensorimotor adaptation. However, I have several comments and suggestions that could improve the paper. Specific comments/questions are listed below:

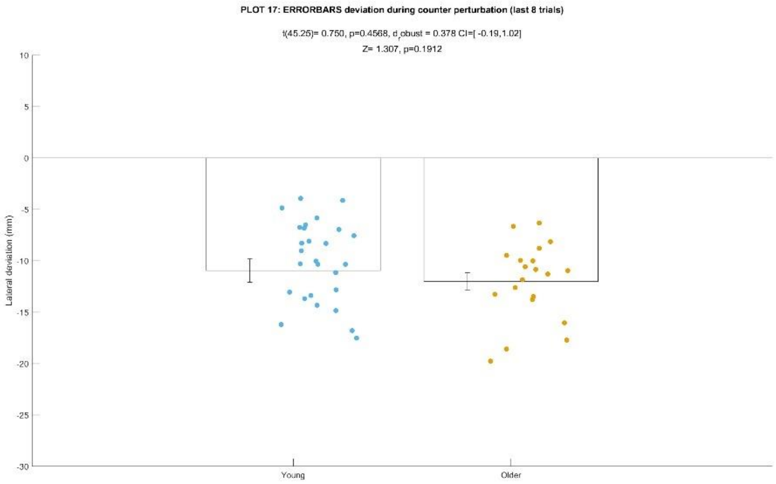
Specific Comments

1. My primary concern is that framing this paper as a replication of Trewartha et al., 2014 is inaccurate. The authors made several methodological choices in designing the experiment that are inconsistent with the procedures used in the Trewartha et al. paper. Indeed, the authors list several methodological differences in the discussion section to consider as reasons for the discrepancy between the results of the two papers. Several additional

methodological differences are mentioned below. I would argue that it would be better to lead this paper with a focus on the more novel questions of A) measuring implicit adaptation in older adults in a force-field adaptation task, and B) evaluating the relationship between implicit adaptation levels and spontaneous recovery. That reframing does not take away from the observation that the spontaneous rebound did not differ in younger and older adults, in contrast to Trewartha et al. That discrepancy warrants further investigation.
Response: Our goal was to disentangle three contradicting hypotheses (spontaneous recovery measures the implicit component of adaptation as postulated by Mcdougle 2015, the implicit component is not affected by aging as postulated Vandevoorde and Orban de Xivry 2019 and the spontaneous recovery is attenuated in older people as postulated in Trewartha et al. 2014). We don't see any valid reasons to attenuate one of these statements in favor of the other two. I believe that this is a subjective matter.

2. It is unclear from the current analyses whether the groups performed similarly during the de-adaptation or washout phase. I did not see a statistical comparison of this phase between groups, although in Figure 2, it looks similar between groups. Given that it is well established that older adults are more susceptible to interference in a variety of learning and memory contexts it would be helpful to show that lateral deviation changed similarly over the de-adaptation trials in the younger and older adults. This would confirm that the washout of prior learning was similar in the two groups.

Response: Thank you for this comment. We did not have any a-priori hypothesis on the role of aging on this very short period of the learning. When we analyzed the last 8 trials of the deadaptation, we, again did not find any significant differences.



We added this information to the paper on line 238-239:
"Total adaptation level at the end of the perturbation phase was similar across age groups (Figure 2B, mean ± SD, young: 2.23±1.43mm, older: 2.29±1.87mm, Analysis 1: t(34.07)= -0.13, p=0.89, Cohen's d = -0.011 CI=[-0.64,0.62]) and at the end of the deadaptation period (trials 298 to 305, t(45.25)=0.75, p=0.46, Cohen's d=0.378 CI=[-0.19, 1.02])."

3. The authors used the last 48 trials of the error-clamp phase to evaluate the spontaneous rebound in the current study. It was not clear to me why they chose not to include the first

16 trials. It is important to note that the largest age differences in spontaneous rebound in the Trewartha et al., paper was from trial 5 through 17 of the error-clamp phase, with several trials after that being statistically similar between groups. This does not seem like an applesto-apples comparison. The current paper also used a longer error-clamp phase. This would provide more time for the younger adults' spontaneous rebound to return towards zero, and potentially, towards the older adults' level.

Response: Trewartha et al. 2014 reports data about spontaneous recovery of motor memories. This is best captured during the later phase of the spontaneous recovery. The earlier phase does not reflect the implicit adaptation that we are interested in. We performed the same statistical analysis as in Trewartha and found no difference on a trial-by-trial basis (which you would off course expect given that single trial data are more noisy that aggregated data)). The p-values for the 64 trials are:

0.843838684363086	0.987683773722990	0.832779909140168
0.500273857120743	0.632168910971130	0.411020702969841
0.264268376730047	0.286334513214071	0.710018284623082
0.830225075820011	0.387509298242931	0.812806979808052
0.341969800396142	0.0315004231901279	0.0489836160824985
0.101345857041200	0.165926050474935	0.482107112786686
0.325984669339512	0.140392112520878	0.302985028908621
0.397500515941385	0.242387193704720	0.729556336863120
0.923301404728378	0.984041225299603	0.585514814688337
0.339831292860710	0.418919271310549	0.508780316103017
0.286079106449480	0.969535229748612	0.317960834222664
0.150614580618972	0.817064387852827	0.733262781231548
0.480045469926394	0.523773817245908	0.849165523096689
0.811893361923770	0.796162528454451	0.520521937375077
0.792350407478056	0.464325798922288	0.921263863169446
0.629962467784511	0.984146908714967	0.732556841367061
0.771278095795972	0.363023393760093	0.801876113099505
0.584690466435748	0.0335628430875101	0.337245506077194
0.824025493501361	0.817251707651819	0.829910847043279
0.911449984276808	0.442274247273209	0.547185194205481
0.655257799099881	0.812480633388200	0.742320032870470
0.415124652296293		

Three of those are <0.05 but, correcting for the multiple comparisons, none of them are $< 0.05/64$.

Our results are also insensitive to the choice of the trial number. We can take all trials from the spontaneous recovery period and obtain the same results:

$t(25.44)= -0.367$, $p=0.72$, Cohen's $d = 0.003$ CI=[-0.58,0.71]

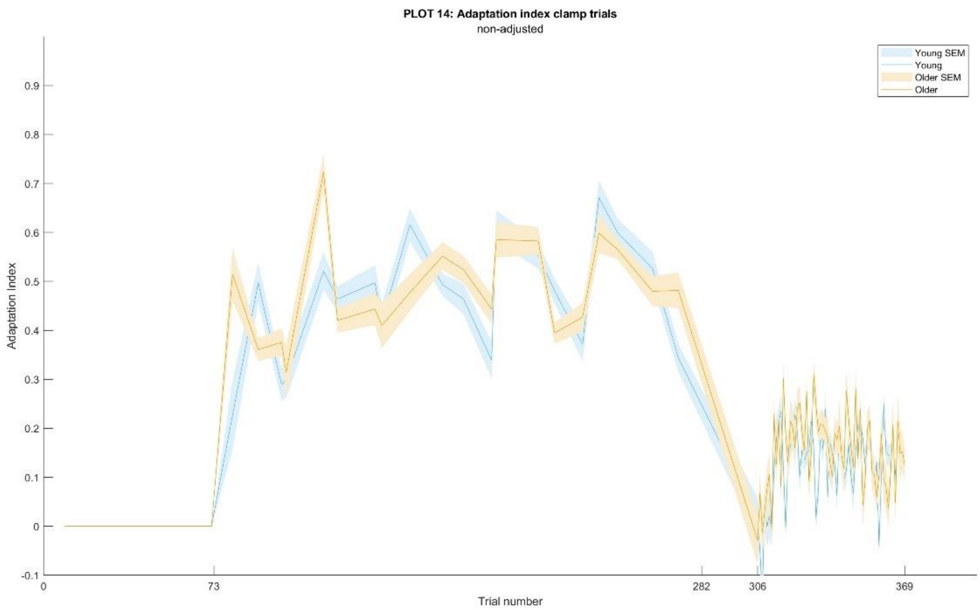
We added all this information to the paper on lines 288-291:

"Note that this result is independent of which trials are analyzed. Performing a trial-by-trial analysis as in Trewartha et al, no between-group differences remained significant after correction for multiple comparisons (p<0.05/64). Similarly, analyzing all 64 trials from the spontaneous recovery period provided the same statistical results (t(25.44)= -0.367, p=0.72, Cohen's d = 0.003 CI=[-0.58,0.71])."

4. Related to the previous comment, it is not clear why the authors did not provide a plot of the adaptation index data for the error-clamp trials to show the spontaneous rebound of the two

groups in a way that aligns with the Trewartha paper. Instead, the exerted force is plotted. Although the exerted force appears consistent with the conclusion that older adults' spontaneous recovery was similar to that of younger adults in the current data, it is not directly comparable to the previous paper. It is also worth noting that the variability within the older adults appears to be much larger than the younger adults. Perhaps there are some older adults with comparable rebound, and some without. There are clearly a handful of older adults exerting even more force than younger adults (Figure 3B). Even looking at Figure 3C, there are several older adults with an adaptation index (albeit in the last 48 trials) below zero, suggesting no rebound, whereas there is only 1 younger adult in the same category. The reason for these individual differences may be unclear but are worth discussing.

Response: In our opinion, the adaptation index data is not a good measure of motor adaptation in force field in the case of slicing movements because the max amount of force is not always aligned on the peak velocity. In addition, this measure contains additional degrees of freedom compared to the force at peak velocity that is uniquely defined (e.g. which portion of the movement do you take into account to compute the adaptation index). The adaptation index can be seen in the figure below.



5. In Figure 4 also, the relatively large forces exerted by 3 of the older adults may contribute to the strength of the correlation. Would the correlation change if those somewhat unusual participants were not included?

Response: these participants are actually deteriorating the correlation. This is even stronger without them. Without those 3 participants, the multilevel correlation is $R=0.61$ (CI: [0.38, 0.77]), $t(43)=5.04$, $p<0.001$

6. I was also curious if there was a correlation between force exerted into the channel and velocity. There appear to be several younger adults with lower velocity relative to the older group. If correlated, it might show that the rebound in younger adults was somewhat lower in this study. Again, it is difficult to evaluate by comparison with the Trewartha paper without a plot of the adaptation index scores across the entire error-clamp phase.

Response: This is a velocity-dependent force-field so the force will be correlated with the velocity. To take this into account, we performed several ANCOVA analyses where we control potential differences in velocity. This is a more powerful approach than normalization (e.g. Stanley 2022, https://doi.org/10.1111/bph.15855).

7. While the authors identify several potential reasons for the discrepancy between their findings and the previous study, the discussion is somewhat dismissive of those potential explanations. Although ample behavioral and statistical evidence is provided from the current sample in favor of the conclusion that older adults' implicit learning is similar to younger adults, the discussion should emphasize the need for additional work to identify the conditions under which older adults might exhibit a smaller rebound than younger adults. This is important in establishing the nature of age-related changes in motor learning more generally. It is worth noting, that some of the factors acknowledged in this paper are not trivial. For example, the length of the adaptation phase may improve the overall memory of the load in older adults. The cognitive aging literature has frequently discussed the role of providing more time for learning in older adults as a way to improve memory performance in various contexts. Individual differences in cognitive abilities, and the potential differences between the respective samples in these studies, should also not be dismissed. It is well known that aging comes with increased interindividual performance variability in many cognitive and motor tasks.

   Response: It is impossible to know why the two studies provided different results. Importantly, as we highlight in the paper, given other observations (implicit adaptation unimpaired in older adults and implicit adaptation reflected in spontaneous recovery), the idea that spontaneous recovery is impaired in older adults does not fit in the big picture. Yes, there are many differences between the two experiments but this result would need to be replicated by the original authors before we can actually build on that result.

8. There appear to be several other methodological differences between the studies that could contribute to the discrepancy in findings. Again, these are not trivial differences. They include the following:

   a. The current older adult sample appears to be younger, and with a tighter range than Trewartha et al. Recent studies in the aging literature have tended towards older age ranges, with mean age often over 70 years old. It is possible that change in the rebound occur after the 60s. Perhaps an exploration of a correlation between chronological age and the magnitude of the rebound would be informative here.

      Response: our data are not powered to look at correlation with age within the older adults. One needs a much larger sample size to do so. In addition, our age range is too narrow to look at this.

   b. The experimental design of the forcefield task here used 8 radial targets from a central start position. The Trewartha study used alternating movements between two targets (akin to Smith et al., 2006) with forces only applied to movements in one direction. Is it possible that implicit learning is strengthened by learning to apply compensatory forces in multiple directions in older adults? To my knowledge, there have not been any studies to directly compare these conditions in older adults. If increasing the number of targets increases the complexity of the task for older adults, it may impact how they perform, an effect commonly observed in a variety of motor tasks.

Response: this is indeed a further difference that we have now added to the manuscript. The reviewer makes an assumption about the effect of target number on age-related differences in adaptation, but this seems an ad hoc assumption.
On lines 427-430, we added:
"The experimental design of the forcefield task here used 8 radial targets from a central start position. In contrast, the Trewartha study used alternating movements between two targets, with forces only applied to movements in one direction. The impact of target number on age-related differences in implicit motor adaptation (or absence thereof) remains unknown."

c. While the cuing method used to quantify implicit learning during the adaptation phase is an important part of this study, it is also a methodological difference from Trewartha et al. that may have impacted how older adults performed. Do those trials provide older adults with a cue about how their movements are adapting over time, and could that impact their level of implicit adaptation and rebound?

Response: this fact was already mentioned in the manuscript, but the interpretation is not as there is absolutely no data to back this up.
We added a sentence to highlight that this might influence the study outcome (line 451):
"We measure implicit adaptation within our adaptation paradigm during learning and working memory in a separate task. This test of implicit adaptation could also have influenced the outcomes of the study."

d. In the current study the deadaptation phase was longer (more trials) than the previous study. This could have an impact if the age groups are differentially impacted by interference (related to my previous comment/question).

Response: This is a direct consequence of the increased number of targets. We have piloted this number to obtain enough de-adaptation in order to be able to observe spontaneous recovery. This difference has been added to line 431: "deadaptation phase (24 vs. 15)"

e. The task instructions for participants in the current study differed from the Trewartha et al. paper. Here, as is common in these experiments, participants were asked to make "slicing" movements through the target and avoid movement corrections. In the Trewartha paper they were asked to stop on the target as the target would become the start position for the next movement. This could have made a critical difference in how the older adults produced the movements in the channel trials.

Response: This makes the task easier and explains why the adaptation index is not a good measure for this study. We added this information to line 437: "While our participants had to slice through the target, those from Trewartha et al. had to stop on the target."

9. The methodological differences noted above are important, especially if it is claimed that this study is an attempted replication of Trewartha et al (2014). Critically, whether this study is a replication or not does not take away from the importance of the current observations or the discrepancy from those previous observations. It is important to highlight those differences to motivate future studies that can explain conditions under which these different findings might be observed. In some ways, focusing so much on framing the paper as a replication, without actually replicating the methods of that previous work, detracts from the potential impact of this paper.

   Response: as outlined in the abstract, we performed a conceptual replication. This means that, in contrast to a direct replication, we copied the scientific hypothesis used in the earlier study in an effort to determine whether the results will generalize to different samples, times, or situations.

   To clarify this further, we added a paragraph on the meaning of a conceptual replication and its consequences on lines 453-459:
   "For all these reasons, our study represents a conceptual replication of the study by Trewartha et al. 2014 and not a direct/exact replication. If any of the factors identified as differences between our study and that of Trewartha is responsible for the difference in outcomes, it means that the age-related effect on spontaneous recovery, if it exists, is highly sensitive to the experimental conditions. By employing multiple methodologies, conceptual replications provide a robustness test of the findings. Our study suggests that the generalizability and robustness of the original results should be considered with caution. It might be true but is likely dependent on the experimental conditions."
   Furthermore, we made sure that we never talk simply about replication but always about conceptual replication.

Reviewer 2

The manuscript presents a comprehensive report on a behavioral study that investigates the effect of age on the contribution of implicit processes to motor adaptation and spontaneous recovery. The introduction presents relevant background information and the data-analysis approach via both frequentist and Bayesian models is an asset to the work. I also appreciate that the authors present individual data points in the figures and I believe that the discussion offers a sound, integrative interpretation of the findings. At the same time, I have several requests/questions related to clarification of the rationale behind the work, the specificity of the hypotheses, the description of the experimental task, and some of the analysis choices. I will provide more detailed comments below, which I hope will be helpful towards revising the work.

1. The introduction states that it currently is unknown whether age affects the implicit component of motor adaptation in a force-field paradigm. It would be helpful for readers if the authors could explain why it would be relevant/important to find an answer to this question. What is fundamentally different about motor adaptation in a force field versus visuomotor rotation paradigm that warrants the investigation performed in the current study? To be clear: I do agree that this is important, but just think it should be clarified more explicitly why this is the case.

Response: We agree with the reviewer that this is important and believe that it is important to know whether previous finding on the absence of age-related differences in implicit motor adaptation true is beyond visuomotor rotation. We added the following sentence in the manuscript (line 50-51):

"Measuring the explicit and implicit components of motor adaptation is essential in order to gain insight into the source of possible deficits."

2. I found the description of the timescales of different processes involved in spontaneous recovery relatively difficult to follow, and am afraid I was lost at the sentence "…which is not washed out by the few deadaptation trials whilst also rapidly forgetting". The sentence seems to involve different theoretical constructs and grammatical structures. Would it be possible to present this in a different manner, such that the contributions of slow and fast processes are addressed separately for example?

Response: we have clarified this by modifying this paragraph. It now reads (lines 64-67):
"*In this framework, the spontaneous recovery of motor memory of field A is attributed to the memory trace of the slow learning process, (McDougle et al. 2015; Smith et al. 2006). This memory trace is masked by the fast adaptation process during the deadaptation period.*"

3. When presenting the hypotheses, please clarify the directionality of the expected group differences and associations. I also recommend adding the research question / hypothesis related to working memory capacity here, as that construct is now only introduced in the methods section, yet it forms an important element in the results and discussion.

Response: We have clarified the directionality of hypothesis 2 ("*spontaneous recovery is larger in young than in older participants*"). We added a fourth hypothesis about working memory that reads "4), as suggested by Trewartha and colleagues, spontaneous recovery is related to explicit memory processes such as working memory capacity.". These changes are on lines 83-87.

4. Have the authors performed a sample size calculation or could they provide another type of justification for including 21-28 participants per group in their experiment?

Response: we initially planned to include 20 participants per group as in Trewartha et al. 2014. We included 20 older participants (recruited 21 but one was excluded due to error in the experiment) and 21 young ones. Then, we noticed that, on average, younger people moved faster than older people in this paradigm despite the speed constraints. We recruited an additional 7 young participants that were instructed to move slower to match hand velocity across age groups (as described on lines 116-118).

We have now added this information on lines 102-107: "*Sample size was initially planned to reproduce the 20 participants per group as in Trewartha et al. 2014. We first included 20 older participants (recruited 21 but one was excluded due to error in the experiment) and 21 young ones. Upon data analysis, we noticed that, on average, younger people moved faster than older people in this paradigm despite the speed constraints (see below). We recruited an additional 7 young participants that were instructed to move slower to match hand velocity across age groups (as described below in the experimental paradigm).*"

5. What were the exclusion criteria for participation related to the general health and consumption habits questionnaires that were used during screening?

Response: The general health questionnaire allows us to make sure that the medical history of the participants do not exhibit that could affect movement control (e.g. head trauma, etc.). The consumption habits questionnaire allows us to look at alcohol/caffeine/drug consumption.

Recreational drug consumption was an exclusion criteria. We added the following (line 94-98): "*Based on the general health questionnaire, participants with events, diseases or injuries that could affect the control of movement were excluded (e.g. head trauma). Based on the consumption habits, participants using recreational drugs or having hazardous alcohol consumption (more than 21 drinks per week for men or more than 14 for women) were excluded from the study. No participants were excluded for these reasons.*"

6. I have several questions about the description of the experimental paradigm;

1. My assumption would be that participants could not just move the handle in the horizontal direction, but also diagonal and forward/backward, as otherwise they would be unable to reach each of the target locations. I therefore wonder is perhaps the statement in line 93 is incorrect? Or do the authors mean to say that the handle moves in 2D?

Response: we indeed mean that the participants moved their hand in the 2D plane. We adapted the sentence as follow (line 109):
"*Participants made center-out, reaching movements in the horizontal plane while holding a robotic handle*"

2. Did the robotic handle return to the starting point after each trial, or did different trial start out from different hand/arm positions?

Response: The participants started from different hand positions and needed to go back to the starting position at the start of each trial. This has been added on line 121: "*and the participants had to move their hand back to the starting position.*"

3. To what extent can the uncued trials really be considered as uncued? Unless I misunderstood the description of the task (in which case, please revise the text as other readers may understand the text similarly), it seems that the white color of the cursor still provides information to participants about what type of trial is to come.

Response: Similarly to Morehead 2015, we referred to cued and uncued trials where uncued trials have a white cursor while the cued trials have a red cursor. We agree with the reviewer that the absence of the red color could also be considered as a cue. We modified the text accordingly (line 137-138): "*While the white cursor could be considered as a cue, we decided to adopt the terminology of Morehead and colleagues (2015).*"

4. With respect to the error-clamp phase, I think it would be helpful to briefly explain what these trials are and whether any feedback was provided. I also strongly recommend to clarify in the text that these trials were not presented on all target locations (currently this is only mentioned in the figure caption).

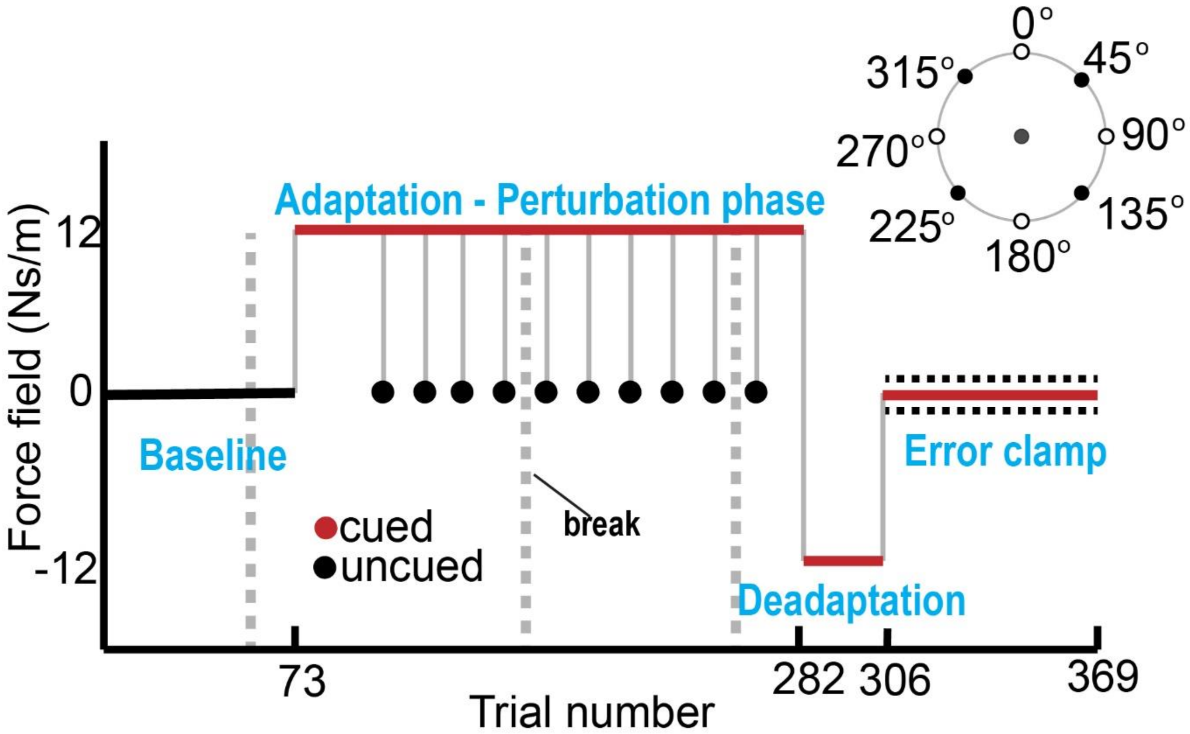
Response: These trials are explained in line 142-144. As done in Smith et al. 2006, visual feedback is provided during these trials (added on line 143). While all 8 targets were used during the retention phase, only 4 of them were used for the error-clamp trials during the baseline and perturbation phase (this has been added on line 145-147).

5. I am not quite sure I understand what it means that there was a channel and the hand was guided through it in the error-clamp phase. Did a physical channel appear in the apparatus?

Response: during a channel, the robot constraints the hand path to a straight line as if it was moving between two walls. We rephrase it to make it clearer (line 142-143): "*by guiding the handle between* *two stiff virtual walls*" We also added a reference to the paper where this technique was first used (Scheidt et al. 2000)

It would be helpful to add some text-info on the phases (e.g., baseline, washout) in Figure 1, which is currently done for the error-clamp phase but not others.

Response: thank you for your advice, we have now changed Figure 1 accordingly.



7. Regarding the WM task, it would be helpful to clarify early on that it concerns a visual WM measure. My suggestion would be to describe both the task and the outcome measure in a separate WM paragraph, rather than as part of the experimental adaptation paradigm.

Response: Thank you for your advice. We have now added a new title: "Experimental paradigm for the visual working memory task". We kept the description of the outcome measure within the data processing and analysis section.

8. I suggest renaming the data analysis paragraph to 'data processing and analysis'. In addition, I think the authors could consider using same order of information as in the task description; this will make the section much easier to follow.

Response: We have reorganized the data processing and analysis section in order to follow the same order as in the paradigm (field trials first and then error-clamp trials).

9. Given that there are quite a larger number of analyses that are being performed, it would be helpful to clarify how they link to the hypotheses. This will also allow for a more concise, integrated presentation (instead of repeatedly saying 'the same was done for…'). Could the authors also explain what levels are included in their multilevel correlation? I assume it includes age to account for differences between groups, but this is currently not described.

Response: We have fully reorganized this section to make it more concise as advised by the reviewer and we specifically mention which analyses are linked to the different hypotheses. It made us realized that all data analyses were not described in this section and added analyses 8 and 9 about the working memory capacity.

10. I missed a description of the WM correlation analyses, and would like to suggest to include age group in the model as well. Work by Seidler has shown that the association between adaptation and WM capacity is different for younger and older adults. It therefore seems reasonable to analyze the younger and older groups separately and/or include group as a factor. Not doing so could potentially have obscured any group-specific results, and the data patterns in Fig. 6 seem to possibly support this notion (though I simply base this on visual inspection, so am not sure that statistical results will concur).

Response: We added analysis 8 and 9 to describe those analyses. We don't believe to have enough power to test correlation on each group separately but we did as instructed by the reviewer. These correlations came out as non-significant both for young and older people. We prefer not to add those to the text as they are highly unpowered.

On lines 222-227, we added the following:

"In analysis 8, working memory capacity was compared between young and older participants with an independent t-test with unequal variance.

In analysis 9, we investigate the possible association (hypothesis 4) between working memory capacity (from analysis 8) and spontaneous recovery levels (from analysis 4) via multilevel correlation from the correlation package in R (Makowski et al. 2020). The different levels corresponded to the different age groups."

11. The use of Bayesian in addition to frequentist statistics to study spontaneous recovery is an asset to the study (although the rationale behind and difference between analysis 4 and 7 were not fully clear to me). The results of the Bayesian analyses do not strongly support the conclusions in the way that PCI defines this (i.e., BF>20), yet I feel this is adequately addressed in the discussion section where the authors acknowledge that the range of effects even includes the possibility of an opposite age pattern). To further support the presented findings, the authors could also apply a Bayesian approach to analyses that go beyond spontaneous recovery, such as the age effect on adaptation and the association between adaptation and WM capacity. This would subsequently allow for interpretations such as the one currently presented in line 13 of the abstract ("both groups adapted equally well" – assuming this is indeed what the BF will support).

Response: We have added a Bayesian analyses for the three outcomes linked to the adaptative response (lateral deviation, force and implicit adaptation at the end of the adaptation phase). The first two analyses yielded moderate evidence while the last one only provided anecdotal evidence for the absence of differences. We have added in the text as:

Line 244 for the lateral deviation outcome: "The corresponding Bayesian analysis suggested that there was moderate evidence an absence of difference (BF = 0.315)."

Line 252 for the force outcome: "The corresponding Bayesian analysis suggested that there was moderate evidence an absence of difference (BF = 0.315)."

Line 269 for the implicit outcome: "The corresponding Bayesian analysis suggested that there was anecdotal evidence an absence of difference (BF =0.45)."

12. When looking at Figure 2A, it seems that there was no complete washout as the deviation does not return to zero. To what extent could this have affected the (interpretation of) the findings? It seems that this is an important point for the discussion section.

Response: While the washout was not complete, the force at the beginning of the error-clamp period was close to zero for both groups, which allowed us to observe a clear rebound. The theory suggests that the amount of washout is actually irrelevant for the spontaneous recovery as the two processes operate in parallel based on the same error.

13. Line 290 may give the impression that data sets from a prior study and the present study were combined. This is a bit misleading, so I suggest to rephrase this.

Response: we have rephrase to (now line 332): “Combining the data of the original study and of the present conceptual replication favor the null hypothesis.”

14. In the discussion, the authors could address practical/clinical implications of their findings (e.g., with regards to physical therapy/rehabilitation after a fall). In addition, I recommend discussing the generalizability to longer-term memory (i.e., savings), or other forms of adaptation (e.g., rotated feedback in the manual domain, but also potential extensions to the locomotor domain).

Response: I am not sure that there are direct practical/clinical implications for this study. We prefer to avoid overselling these results

15. Lines 447-449; It would be helpful if the authors could explain why the inter-subject variability difference between younger and older adults is a limitation. How does it affect the (boundaries of) interpretation of the present findings? I would also like to read more about how they envisage this issue being tackled. Do the authors mean that researchers should look for more homogeneous older adult groups? Or that specific methodological/statistical approaches should be developed to

Response: The variability of the outcomes makes it harder to estimate whether there is an effect or not and how big this effect is.

16. Throughout the manuscript there are several instances were small corrections or additions to the text would be very helpful for readers. I list some examples from the introduction section here, but encourage the authors to check for other similar issues in the remainder of the manuscript;

1. Lines 23 and 25; I suggest changing the phrasing of ‘thanks to’ to ‘via’ or ‘through’. → we changed them to “via”

2. Line 27; Could the authors please clarify what it is that is being updated? I think they may be referring to motor representations, but this is currently up to the reader to infer and would be easier to understand if it was made explicit. → we change it to updating the movement.

3. Line 36; What do the two sets of references mean? → they should be merged.

4. Line 37; This should read ‘suggests’. → this has been corrected.

5. Line 44; Please explain what the little difference refers to; I think it refers to adaptation performance between younger and older age groups, but this should be clarified. → Correct, it now reads “Little difference in the amount of adaptation to a force-field perturbation has been found between young and older participants”

6. Line 53; ‘which IS hidden’ → this has been corrected.

7. Line 67; I'm not sure how three things can be contradictory. → there is a logical problem here. If implicit adaptation is unimpaired in older people and if it determines spontaneous recovery, then spontaneous recovery cannot be different across age groups. This later sentence has been added to the manuscript on line 73-75.

8. Line 73; Start the sentence with 'As it is known…' → this has been changed.