**Manuscript title:** How movement variability constrains locomotor use-dependent learning

**We thank the reviewer for their comments regarding our power analysis. Below, we clarify exactly which data we used for our power analysis (experiment 2 from Wood et al 2020) and provide a clearer rationale for why we believe these data provide us with a reasonable estimate of effect size. Importantly, we also directly address the reviewer’s overarching concern regarding statistical power: *We now plan to increase our desired sample size by 20%, for a total of 18 participants.* Please note that we have broken up the reviewer’s comment into several parts, so that we could effectively address each point.The reviewer’s comment is copied below in normal font and our responses follow in bold. Pointers to line numbers and figures refer to the tracked changes version manuscript unless otherwise specified.**

My only comment is that the additional information in the power analysis section is confusing. The study was powered using data from the Wood 2020 paper, which introduced the perturbation gradually. Yet in the revision the authors now describe calculating power from "an abrupt learning phase" of an experiment. The only "abrupt phases" that I see in Wood 2020 are well into the learning process when the visual stimulus is turned off and then back on again? So, I am confused.

**We apologize for our lack of clarity in our previous response. As the reviewer notes, our estimate of effect size in the power analysis is based on the step asymmetry index (SAI) differences between the Use-Dependent Plasticity (UDP) and UDP Dose-Response (UDP-DR) groups at the early washout phase of experiment 2 (Fig. 4 from Wood et al 2020. We considered the immediate switch from no feedback during the catch 2 phase to full step asymmetry magnitude during the learning 3 phase an “abrupt learning phase”, and thought it appropriate as it immediately preceded the early washout phase. However, we now see how this was confusing, since the perturbation was gradually introduced at the start of the first learning phase. We have removed the reference to the abrupt learning phase and instead state the exact timepoint from which we made our effect size estimate in the main text (lines 386-392). We further address the reviewer’s comment about the gradual perturbation in the response regarding overestimating power.**

I would have expected the authors to run the power analysis on data from early in the washout phase of the experiment.

**Yes, this is exactly what we did. We used the SAI data from the early washout phase during experiment 2 (Fig. 4 from Wood et al 2020).**

I am left wondering if the power analysis is based on early washout from experiment 2 in the Wood 2020 paper. In that experiment, different magnitude perturbations were given with one twice the size of the other (both introduced gradually).

**Yes, this is exactly correct. The larger of the two perturbations in Wood et al 2020 had a magnitude of 22%, exactly the same mean magnitude we are testing in all three conditions in the current registered report. We clarify this point in the following response.**

If this is the case, I worry that you might be over-estimating the power that you will have to detect a difference.

**We thank the reviewer for their comment, as it has helped us clarify our own thoughts on the statistical power in our registered report, and we hope that our response here directly addresses the reviewer’s concern. With regard to our estimated effect size, the gradual perturbation in Wood et al 2020 is more likely to lead to a conservative estimate of effect size than an overestimation. Concretely, there is no theory that predicts more use-dependent learning from a gradually introduced versus an abrupt perturbation. (This may be another important distinction between use-dependent learning and error-based adaptation.) Our Strategy plus Use-Dependent model predicts no effect of variability, while our Adaptive Bayesian model indicates that gradual perturbations could lead to attenuated use-dependent learning relative to the abrupt perturbation we currently propose. During a gradual perturbation, participants must execute a wide range of step lengths to reach the desired 22% SAI. In contrast, participants quickly adjust their SAI to an abrupt perturbation, as our pilot data show (Fig. 4), yielding less overall movement variability. According to the Bayesian model, increased variability—as in the gradual case—delays development of a use-dependent bias.**

**As far as the use of perturbations of different magnitudes in Wood et al 2020, here we propose to employ a mean target SAI of 22% during learning for each condition and experimentally manipulate perturbation variability, rather than magnitude. Despite this difference in methods between studies, we believe these data still provide us with a reasonable estimate of potential effect sizes. As a result of the reviewer’s comment, we simulated the Adaptive Bayesian model using parameter values obtained by fitting individual participant data from Wood et al 2020 (lines 449-451). The predicted effect sizes are modestly larger (dz = 0.95 - 1.01) than the one from Wood et al 2020 used in our power analysis (Cohen’s d = 0.91).**

**A final important point we did not make clear in our initial response was that Wood et al 2020 was a between-subjects design with n=10 and 16 participants in the UDP and UDP-DR groups, respectively. Here, we now plan to recruit 18 participants in a within-subjects design, which should increase statistical power.**

**Regardless of these rationales, we also recognize there is inherent uncertainty in all power analyses. *Therefore, we believe the best safeguard against an underpowered study is to increase our sample size, which is why we are increasing our desired sample size by 20%.***

I bring this up because I expect that the effects in the planned experiment will be rather small (based on your simulations).

**We appreciate the reviewer’s concern, which has led to our decision to test more participants. Based on our simulations, use of a within-subjects design, and now larger sample size, we believe we should be appropriately powered. Also, while the raw aftereffect sizes from UDP appear small, they are robust. All 16 participants from the UDP DR group in Wood et al demonstrated use-dependent biases, underlying the very large within-subjects standardized effect size we have since calculated for this group when comparing early washout relative to late baseline (dz=1.94).**

It would be helpful for you to clarify exactly what data were used for the power analysis.

**Thank you for your helpful comments. We hope it is now clear exactly which data we used in the power analysis, and that we have satisfactorily addressed all remaining concerns. For convenience, the changes to the main text are copied below, with tracked changes:**

***We performed a power analysis to determine the sample size required to detect differences in use-dependent biases across conditions, with alpha of 0.05 and power of 0.90. We estimated a standardized effect size (Cohen’s d) of 0.91 using group step asymmetry differences during the early washout phase of experiment 2 from Wood et al., 2020. Based on this estimated effect size, we will require 15 subjects. However, in order to ensure we safely exceed this threshold for power, we plan to recruit 18 subjects. We therefore expect to recruit 18-24 individuals for this study in order to account for possible attrition and to exceed the minimum acceptable power. This sample size will also ensure appropriate counterbalancing of practice schedules across participants while also being well-above the threshold of statistical power documented in comparable motor learning studies (Diedrichsen et al., 2010; French et al., 2018; Long et al., 2016; Verstynen and Sabes, 2011; Wood et al., 2020).***