We thank both reviewers for their time and effort reviewing our manuscript. We appreciate their thoughtful comments intended to improve the scholarly quality of this paper. As reviewer 2 had only one request, which was also made by reviewer 1, we ask that reviewer 2 direct their attention to the end of this document, which addresses both reviewers' request to add a figure to the manuscript.

Our responses to comments from reviewer 1 (in bold) follow.

Methods:

1. Definition of MA side – an objective metric would have been preferred, even though they are not without their limitations.

The definition of MA side used in our paper (i.e. side of first PD symptoms) is an unambiguous classification which has been used previously by our and other research groups (Sofuwa et al., 2005; Louie et al., 2009; Roemmich et al., 2013, 2014; Siragy et al., 2020). Furthermore, while the MA side may change during the progression of PD symptoms, the side of onset typically remains the more affected side (Winogrodzka et al., 2005; Erra et al., 2019). Evidence suggests that the side of first symptomatic emergence affects the nature and course of PD (Munhoz et al., 2013; Poletti et al., 2013), and that handedness is a key predictor of the onset side (Uitti et al., 2005). Therefore, although other definitions have merit—albeit with different implications—we chose the side of onset as the MA side based on this literature.

Kindly clarify – was the definition made separately for upper and lower limbs?

The definition of MA side was not specific to the upper or lower limbs. The determination of MA side was the side of first symptomatic emergence, regardless of upper or lower limb.

Kindly consider/discuss – How robust do you think the self-report was (in terms of recall bias), and based on the UPDRS III, were these sides still the most affected?

We are confident in the reliability of self-reported side of symptomatic emergence: the clinical assessment and determination of MA side were conducted in the presence of participant's caregivers/partners, who verified the accuracy of MA side reporting. Self-reporting the more affected side is common in studies of Parkinson's disease (Sofuwa et al., 2005; Louie et al., 2009; Roemmich et al., 2014).

2. Task protocol – The description of the methods could be more elaborate

Kindly clarify – How was the preferred walking speed determined and standardized?

The standard method used to determine preferred walking speed was by adjusting treadmill speed up and down until participant's affirmed their comfort with the current speed.

Kindly clarify – Was the background containing other stimuli apart from the words? Was the challenge to recognize words in the noise? How was the contrast determined? Was the duration/frequency of the words sufficient to demand continuous attention to the task? 80 + 20 seconds is less than 2 minutes – was there no dual task for the last 20 seconds of the dual task trial?

The projection screen of the CAREN system was showing a virtual park path, that included benches, trees, and shrubbery along the path (Sinitski et al., 2015). The task was simply to require active spatial awareness, similar to that required in daily life, such as looking for a particular gate at an airport. Therefore, the words were in a clearly visible font/contrast. The reviewer is correct; the first and last 20 sec of the dual-task trial did not have the dual-task. Only the portion of the dual-task trial during which the task words were being shown was analyzed.

3. Outcomes

Kindly clarify – How was DTC calculated?

The following sentence was added to the manuscript: "Dual task cost was defined as the difference between single and dual task for a given variable."

4. Medication state

Kindle discuss - the possible impact of medication state on coordination results, contrasting with the earlier mentioned research

We are unsure which "earlier mentioned research" the reviewer is referring to; there is no discussion of medication currently in the paper or mentioned earlier in the reviewer's comments (either reviewer). There is very limited research on the effects of medication in PD on biomechanical related measures, and even further limited for papers focused on coordination. We are not aware of any studies focused on the effects of medication on coordination in PD related to dual-tasking and gait. Furthermore, as our study did not include ON/OFF medication as an experimental condition, it would be inappropriate to speculate on the impact of the medication state on our measures of coordination. We have revised the concluding sentence as follows:

"More research is needed to support or dispute the existence of differences in coordination between the MA and LA sides, and whether any differences are moderated by medication."

Statistics

5. Equivalence testing – is it applicable and applied correctly?

Kindly reconsider approach – the smallest effect size detectable with the current sample size is d= 0.9, yet the equivalence bounds used are far smaller than that (d=0.36). This seems contrary to the recommendations in the Lakens paper which suggests using the smallest detectable effect size to determine the bounds.

We acknowledge our low power for the chosen bounds. In the cited paper, Lakens also recommended choosing informative bounds when possible (i.e. relevant literature exists), and noted the risk of bias when using standardized effect sizes for equivalence bounds. Clinical relevance (i.e. comparing to estimate based on healthy older adults) was deemed potentially more informative than choosing bounds based on the smallest powered effect size (based on sample size). Equivalence testing is applicable to our research goals, as it gives a (qualified) proof of a null result (i.e. no different, or "equivalent", to a healthy population). This theoretically allows answering the question of whether coordination between MA/LA sides in PD is within a typical range for

healthy older adults, or whether it is pathological. We believe that our complete reporting of power sensitivity analysis, statistical results, and underlying code and data provide readers with enough information to exercise their own judgment, if they have different standards, needs, or use of our results.

Kindly clarify – only one bound is mentioned for the various outcomes, is this the upper bound, and if so, what is the lower bound? Or is inferiority testing being applied, rather than equivalence testing?

The equivalence bounds reported in the Statistics section are symmetric (+/-); this reporting follows common reporting and discussion of equivalence testing and bounds, where singly reported bounds for an "equivalence test" are treated as symmetric unless otherwise noted (i.e. an equivalence test implies symmetric bounds unless specifically otherwise noted). (Additionally, as you suggest, asymmetric bounds, specifically where one bound is treated as infinite, are the basis for inferiority or superiority tests, which are a subset of equivalence testing with asymmetric equivalence bounds.)

6. Multiple comparison problem

Kindly consider – reporting multiple-comparison corrected p-values

There were no multiple-comparisons performed on the same dependent variable. Additionally, we are not making higher level claims based on the collective statistical results of multiple dependent variables, therefore, a false-discovery rate control procedure, such as described by Benjamini & Hochberg (1995), is similarly unnecessary.

Results

7. Demographics

Kindly clarify – For a mid-stage PD population such as this, including freezers and fallers, the UPDRS scores are extremely low. Were these scored by a movement disorders specialist?

JN performed the UPDRS assessment and is a trained UPDRS rater. Participants were tested when optimally medicated, i.e. when they felt their medication was most effective. The sentence on participant medication status in the methods was revised as follows:

"Participants were tested when optimally medicated ("ON") by their normal medication."

Additionally, as we noted in the paper, the UPDRS scores at 10 ± 5 designates this group as mild to moderate PD. While freezing and/or falling is more common in more severe PD, it can begin at any stage of the disease; UPDRS tests include items for freezing and falling, but encompasses a broader set of criteria. As well, our classification of "fallers" is conservatively based on the self-report of having fallen at any time (even once) in the year prior to the time of data collection; this may encompass non-injurious falls, and environmentally-caused falls not reflective of the participant's abilities. We have added the following sentence to the Methods:

"Falling status was self-reported based on conservative criteria of having fallen at least once in the past year, including non-injurious falls."

8. Presentation of results

Kindly adjust – Bilateral difference in DTC could also go into a table to reduce the text

We have converted the data in that paragraph to a table; please see the revised manuscript for the addition.

9. Presentation of data

Kindly adjust – add figures to show individual data points for important results.

We have added a figure to bolster our discussion point that the presence or absence of significant differences in either side does not suggest the presence or absence of significant differences between the sides. We have also clarified some phrasing that was ambiguous about which variable we were referring to in that paragraph. Please see the revised manuscript for the figure and updated paragraph.

References:

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, **57**, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Erra C, Mileti I, Germanotta M, Petracca M, Imbimbo I, De Biase A, Rossi S, Ricciardi D, Pacilli A, Di Sipio E, Palermo E, Bentivoglio AR, Padua L (2019) Immediate effects of rhythmic auditory stimulation on gait kinematics in Parkinson's disease ON/OFF medication. *Clinical Neurophysiology*, **130**, 1789–1797. https://doi.org/10.1016/j.clinph.2019.07.013

Louie S, Koop MM, Frenklach A, Bronte-Stewart H (2009) Quantitative lateralized measures of bradykinesia at different stages of Parkinson's disease: The role of the less affected side. *Movement Disorders*, **24**, 1991–1997. https://doi.org/10.1002/mds.22741

Munhoz RP, Espay AJ, Morgante F, Li J-Y, Teive HA, Dunn E, Gallin E, Litvan I (2013) Long-duration Parkinson's disease: Role of lateralization of motor features. *Parkinsonism & Related Disorders*, **19**, 77–80. https://doi.org/10.1016/j.parkreldis.2012.07.008

Poletti M, Frosini D, Pagni C, Baldacci F, Giuntini M, Mazzucchi S, Tognoni G, Lucetti C, Dotto PD, Ceravolo R, Bonuccelli U (2013) The relationship between motor symptom lateralization and cognitive performance in newly diagnosed drug-naïve patients with Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology*, **35**, 124–131. https://doi.org/10.1080/13803395.2012.751966

Roemmich RT, Field AM, Elrod JM, Stegemöller EL, Okun MS, Hass CJ (2013) Interlimb coordination is impaired during walking in persons with Parkinson's disease. *Clinical Biomechanics*, **28**, 93–97. https://doi.org/10.1016/j.clinbiomech.2012.09.005

Roemmich RT, Nocera JR, Stegemöller EL, Hassan A, Okun MS, Hass CJ (2014) Locomotor adaptation and locomotor adaptive learning in Parkinson's disease and normal aging. *Clinical Neurophysiology*, **125**, 313–319. https://doi.org/10.1016/j.clinph.2013.07.003

Sinitski EH, Lemaire ED, Baddour N, Besemann M, Dudek NL, Hebert JS (2015) Fixed and self-paced treadmill walking for able-bodied and transtibial amputees in a multi-terrain virtual environment. *Gait & Posture*, **41**, 568–573. https://doi.org/10.1016/j.gaitpost.2014.12.016

Siragy T, MacDonald M-E, Nantel J (2020) Restricted Arm Swing in People With Parkinson's Disease Decreases Step Length and Time on Destabilizing Surfaces. *Frontiers in Neurology*, **11**.

Sofuwa O, Nieuwboer A, Desloovere K, Willems A-M, Chavret F, Jonkers I (2005) Quantitative Gait Analysis in Parkinson's Disease: Comparison With a Healthy Control Group. *Archives of Physical Medicine and Rehabilitation*, **86**, 1007–1013. https://doi.org/10.1016/j.apmr.2004.08.012

Uitti RJ, Baba Y, Wszolek ZK, Putzke DJ (2005) Defining the Parkinson's disease phenotype: initial symptoms and baseline characteristics in a clinical cohort. *Parkinsonism & Related Disorders*, **11**, 139–145. https://doi.org/10.1016/j.parkreldis.2004.10.007

Winogrodzka A, Wagenaar RC, Booij J, Wolters EC (2005) Rigidity and bradykinesia reduce interlimb coordination in Parkinsonian gait. *Archives of Physical Medicine and Rehabilitation*, **86**, 183–189. https://doi.org/10.1016/j.apmr.2004.09.010